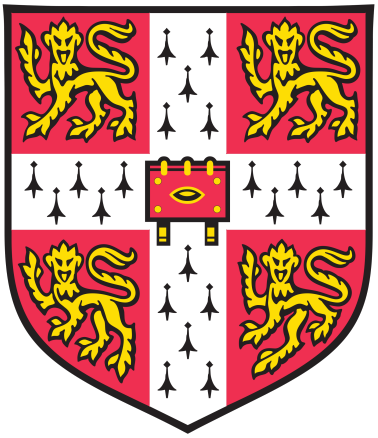
Zonal flows in accretion discs and their role in gravito-turbulence



Riccardo Vanon

Department of Applied Mathematics and Theoretical Physics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

St Edmund's College

April 2017

Zonal flows in accretion discs and their role in gravito-turbulence

Riccardo Vanon

Abstract

This thesis focuses on the evolution of zonal flows in self-gravitating accretion discs and their resulting effect on disc stability; it also studies the process of disc gravito-turbulence, with particular emphasis given to the way the turbulent state is able to extract energy from the background flow and sustain itself by means of a feedback.

Chapters 1 and 2 provide an overview of systems involving accretion discs and a theoretical introduction to the theory of accretion discs, along with potential methods of angular momentum transport to explain the observed accretion rates. To address the issue of the gravito-turbulence self-sustenance, a compressible non-linear spectral code (dubbed `CASPER`) was developed from scratch in `C`; its equations and specifications are laid out in Chapter 3.

In Chapter 4 an ideal (no viscosities or cooling) linear stability analysis to non-axisymmetric perturbations is carried out when a zonal flow is present in the flow. This yields two instabilities: a Kelvin-Helmholtz instability (active only if the zonal flow wavelength is sufficiently small) and one driven by self-gravity.

A stability analysis of the zonal flow itself is carried out in Chapter 5 by means of an axisymmetric linear analysis, using non-ideal conditions. This considers instability due to both density wave modes (which give rise to overstability) and slow modes (which result in thermal or viscous instability) and, thanks a different perturbation wavelength regime, represents an extension to the classical theory of thermal and viscous instabilities. The slow mode instability is found to be aided by high Prandtl numbers and adiabatic index γ values, while quenched by fast cooling. The overstability is likewise stabilised by fast cooling, and occurs in a non-self-gravitational regime only if $\gamma \lesssim 1.305$.

Lastly, Chapter 6 illustrates the results of the non-linear simulations carried out using the `CASPER` code. Here the system settles into a state of gravito-turbulence, which appears to be linked to a spontaneously-developing zonal flow. Results show that this zonal flow is driven by the slow mode instability discussed in Chapter 5, and that the presence of zonal flows triggers a non-axisymmetric instability, as seen in Chapter 4. The role of the latter is to constrain the zonal flow amplitude, with the resulting zonal flow disruption providing a generation of shearing waves which permits the self-sustenance of the turbulent state.

Dedication

To Mum and Dad,
for their unrelenting love and for always believing in me.

To my girlfriend Lyn,
for never failing to cheer me up in the difficult moments.

Acknowledgements

The writing of this doctoral thesis has only been possible thanks to the massive amount of support I have received during the duration of my PhD.

Above all, the emotional support of my parents, girlfriend, sister, grandmother, aunt and uncle has been invaluable when going through difficult moments in my doctoral journey.

Similarly, the technical help I received from my supervisor, Prof. Gordon Ogilvie, has been extremely beneficial. He has been incredibly patient and helpful throughout my PhD journey, making this thesis possible. I would also like to thank Dr. Henrik Latter, John Ross and Dr. Adrian Barker for their help and for useful discussions, as well as the rest of DAMTP for the fantastic and motivating atmosphere they provided during my time at the Department.

Last but not least, a heartfelt thank you to all those who ever believed I could reach for the stars: my friends (spread in quite a few countries!) and my teachers – in particular Professoressa Sabbatini and Mr. Okon – who encouraged me to keep aiming higher.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

Riccardo Vanon
April 2017

Contents

List of Tables

Page

List of Figures

Page

List of Acronyms

AGN . . . Active Galactic Nuclei

AU . . . . Astronomical Unit

BH . . . . Black Hole

CFL . . . Courant-Friedrichs-Lewy

CR . . . . Co-rotation

CV . . . . Cataclysmic Variable

EUV . . . Extreme Ultra-Violet

FD . . . . Finite Difference

FE . . . . Finite Element

FUV . . . Far Ultra-Violet

GI . . . . Gravitational Instability

GRB . . . Gamma Ray Burst

GSF . . . Goldreich-Schubert-Fricke

HMXB . High-Mass X-ray Binary

IC . . . . . Initial Condition

ILR . . . . Inner Lindblad Resonance

IR . . . . Infrared

KH . . . . Kelvin-Helmholtz

LMXB . . Low-Mass X-ray Binary

MHD . . Magnetohydrodynamics

MRI . . . Magnetorotational Instability

NS . . . . Neutron Star

ODE . . . Ordinary Differential Equation

OLR . . . Outer Lindblad Resonance

PDE . . . Partial Differential Equation

PP . . . . Proto-planetary

RWI . . . Rossby Wave Instability

SBI . . . . Sub-critical Baroclinic Instability

SDW . . . Spiral Density Wave

SED . . . Spectral Energy Distribution

SG . . . . Self-gravity

SMBH . . Super-Massive Black Hole

VSI . . . . Vertical Shear Instability

WD . . . . White Dwarf

XB . . . . X-ray Binary

YSO . . . Young Stellar Object

Chapter 1

Systems with accretion discs

This thesis focuses on the presence and evolution of zonal flows in accretion discs and the role they play in disc stability.

In order to establish this, three different methods of analysis are employed. Two different linear analyses of an accretion disc are carried out that involve zonal flows. In the first approach, the survival of zonal flows is investigated by imposing an axisymmetric zonal flow on the disc and studying its dynamical stability with respect to non-axisymmetric perturbations. In the second approach, the onset of zonal flows is investigated by studying the stability of a homogeneous disc with respect to the development of axisymmetric perturbations, including zonal flows, in the presence of viscosity, thermal diffusion and radiative cooling. The last approach is to develop a fully non-linear, compressible spectral code to simulate a patch of the self-gravitating disc using the shearing sheet model. The diversity in the approaches taken is justified by their complementing strengths: linear analyses allow a quicker assessment of the problem at hand, permitting to consider the dependence of accretion disc stability on a wide range of parameters and conditions. The linear analyses presented here consider the presence/absence of disc self-gravity, viscous/diffusive effects, cooling and a range of values for the adiabatic index, Toomre parameter, zonal flow wavelength and perturbation wavelength. The non-linear code, on the other hand, offers a more complete picture of the flow's behaviour and allows us to investigate the matter of disc gravito-turbulence and the source of its self-sustaining feedback by drawing on results obtained using the linear analyses.

This Chapter and Chapter 2 present an introductory view of the topic. Because

the work discussed in this thesis is of a general nature, and only at times features conditions (e.g. self-gravity) which may indicate the type(s) of disc/system considered, this chapter describes in a generic and brief way the astrophysical systems which typically host accretion discs. Chapter 2, on the other hand, focuses on the theory of disc dynamics; as accretion can only occur via the outward transport of angular momentum, possible mechanisms allowing this are discussed, with a somewhat greater emphasis placed on self-gravity. Finally, it gives an overview of zonal flows: their occurrence in astrophysical systems as well as in the literature, and some of the instances in which these have been found to play an important role in the dynamics of the system considered.

Accretion discs are common objects in the Universe and they can occur in various types of astrophysical systems. Regardless of this, all accretion discs behave similarly; their ring-like configuration is due to the action of gravity and angular momentum, with particles within the disc rotating around the central object in a differential fashion. This creates a frictional shear force between any two orbits of the disc which allows inner orbits to pass angular momentum to outer orbits; this outward transport of angular momentum (which is dealt with in more detail in Section 2.2) is a very crucial part of the disc dynamics, for it allows matter to be transported inwards instead, eventually leading to accretion onto the central object.

Although the idea of accretion discs has roots in the 18th century with the notion by Kant (1755) that the planets of our Solar System were formed within a thin disc of material previously surrounding the Sun, the basic equations describing a disc's evolution were only derived nearly 200 years later by von Weizsäcker (1948) and Lüst (1952). In 1964, the works by Zeldovich (1964) and Salpeter (1964) first drew attention to the large amount of gravitational energy that is released during accretion, although they did not consider an accretion disc but a simpler case of a compact object accreting interstellar gas. In the following years, groundbreaking works such as Saslaw (1968) and Lynden-Bell (1969) considered accretion from discs as a possible explanation for energetic phenomena such as Cataclysmic Variables (CVs), dwarf novae and quasars.

In the 70s accretion disc theory became steadily more popular in the literature, also thanks to the establishment of a 'standard accretion disc model' by Shakura and Sunyaev (1973) (further examined in Section 2.1.2); this model makes use of simple assumptions to parameterise the viscosity due to turbulent motions – whose

presence enables angular momentum to be transported outwards, hence allowing accretion – in a geometrically thin, optically thick disc.

The model by Shakura and Sunyaev (1973) – as well as other variations such as geometrically thick discs, hot discs and advection dominated accretion flows – provided good agreement with existent observations, making accretion discs a fundamental part of models describing CVs, Active Galactic Nuclei (AGN) and many other types of system. Below is a short outlook on some of the astrophysical systems believed to be linked with the presence of accretion discs.

1.1 Young Stellar Objects

Accretion discs around Young Stellar Objects (YSOs), usually called proto-planetary (PP) discs for they represent the cradle for the formation of planets, form following the collapse of a molecular cloud. In such a cloud thermal pressure opposes the contracting action of gravitational interactions; the cloud is therefore found to be stable if it obeys hydrostatic equilibrium

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{G\rho(r)M_r}{r^2}, \tag{1.1}$$

where P is the pressure, G the gravitational constant, $\rho(r)$ the density of the gas at some radius r and M_r the mass enclosed within the same radius.

The eventual fate of the cloud is determined by the timescales on which the destabilising effect of gravity and the stabilising thermal pressure act. The latter phenomenon is characterised by the sound-crossing time

$$t_{\mathrm{cross}} = \frac{R}{c_{\mathrm{iso}}}, \tag{1.2}$$

where c_{iso} is the isothermal sound speed, which simply represents the time taken for sound waves to cross a cloud of radius R. Gravitational interactions, on the other hand, occur on a free-fall timescale

$$t_{\mathrm{ff}} = \frac{1}{\sqrt{G\rho}}. \tag{1.3}$$

Instability occurs if $t_{\mathrm{ff}} < t_{\mathrm{cross}}$, as in this case pressure effects act too slowly to balance out the gravitational collapse and a runaway contraction is triggered. From Equation 1.3 it is possible to see that an increase in the mass of the cloud at fixed R

shortens the free-fall time, hence destabilising the system as the denser environment enhances the gravitational interactions.

By equating the two timescales, it is possible to find the critical radius R_J – called the Jeans' length – at which the system is in equilibrium. Its form is given by

$$R_J = \frac{c_{\rm iso}}{\sqrt{G\rho}}, \tag{1.4}$$

with clouds larger than the Jeans' length being unstable. A similar stability criterion is represented by the Jeans' mass, which can be simply thought of as the mass enclosed within a cloud having a radius equal to the Jeans' length.

After the YSO forms at the core of the cloud following its gravitational collapse, a disc takes shape around it as material further away from the core, hence possessing a larger angular momentum, falls inward (Williams and Cieza, 2011). The disc's outer edge is delimited by the centrifugal radius, which is however a function of the infall time and the core's angular rotation frequency $R(t) \propto \Omega^2 t^3$ (Terebey et al., 1984; Williams and Cieza, 2011)[1], underlining the sensitivity of the disc's evolution to the core properties and the vast diversity of initial disc properties.

Disc stages

Protoplanetary discs are grouped into four classes depending on the characteristics of their spectral energy distributions (SEDs); it is believed the sequence of classes is in fact correlated to the temporal evolution of the star/disc system. The protostar in a disc of Class 0 is still well embedded in the gaseous envelope and the two components' contributions to the SED are indistinguishable. In Class I systems, defined by a IR spectral slope $b_{\rm IR} = \mathrm{d}\log\lambda_{\rm IR}F_\lambda/\mathrm{d}\log\lambda_{\rm IR} > 0$, the disc and the envelope are believed to have roughly similar mass and both contribute to the SED signature. Accretion from the envelope has all but ceased in systems belonging to Class II, most of the mass now residing in either star or disc; these systems are characterised by $0 > b_{\rm IR} > -3/2$. By the time the system reaches the final Class III, accretion from the disc to the protostar has substantially reduced, leaving a T Tauri star surrounded by a low mass accretion disc, where planet formation is believed to occur; these systems exhibit a spectral slope of $b_{\rm IR} < -3/2$. Class III

[1]Basu (1998) argues that the dynamics of a magnetised collapsing core might be altered by a differential, rather than rigid, rotation; this would cause the centrifugal radius to only increase linearly as a function of time.

is the category with the longest timescale among the four, allowing for a bias in the observations in favour of this Class; also, thanks to the material having either accreted or being blown out by stellar winds, they are unobscured and can be observed with higher resolution, shorter wavelength equipment.

Accretion

Observations show that accretion onto YSOs occurs in episodic outbursts in FU Orionis variables, which generally last in the range of years to decades, producing peak accretion rates of $\dot{M} \sim 10^{-4} M_\odot$ year^{-1} (Hartmann and Kenyon, 1996; Armitage, 2011). The presence of outbursts in the early stages of PP discs is also backed up by statistical evidence: in particular, a discrepancy is seen in the actual bolometric luminosity of the proto-stars and the larger luminosity expected for a steady accretion given the rate of infall and the timescale for the proto-stellar phase (Kenyon et al., 1990).

The outbursts in FU Orionis variables are thought to be triggered by a thermal instability in the disc (also responsible for dwarf nova outbursts, as seen later in Section 1.2.3); a disc is said to be thermally unstable when its equilibrium – characterised by local heating and cooling rates $\mathcal{H}$ and $\mathcal{C}$ balancing each other out – is unstable to small perturbations in its central temperature T_c. Mathematically, this can be written as (Pringle, 1976)

$$\frac{\partial \ln \mathcal{H}}{\partial \ln T_c} > \frac{\partial \ln \mathcal{C}}{\partial \ln T_c}. \tag{1.5}$$

In the eventuality of some disc regions being locally unstable to thermal instability, a limit cycle – in which the disc alternates between quiescent (where its midplane is cool and neutral and the accretion rate to the central star low) and outburst phases (characterised by hot disc temperatures, high ionisation and large accretion rates) – can occur as schematised in Figure 1.1.

This limit cycle can occur once the midplane temperature approaches the rough critical value $T_c \sim 10^4$ K, for in this temperature region the opacity is very sensitive to changes in T_c (Armitage, 2011). This produces multiple solutions in the thermal equilibrium function of the local vertical structure at a given radial distance r from the central star, usually referred to as a "S-curve". This function in the ($\log \Sigma$, $\log T_{\rm eff}$) plane can be de-constructed in three branches, divided by the two inflection points: the stable top and bottom branches (as here the direct correlation of surface

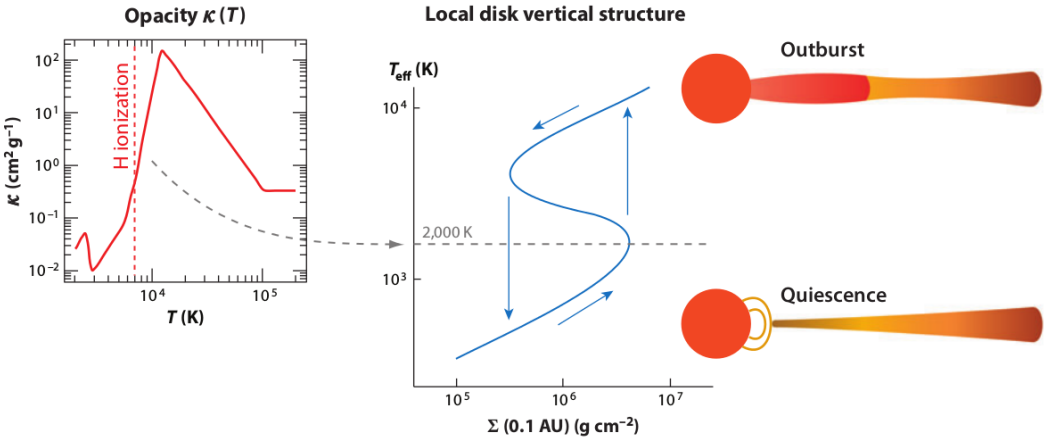


Figure 1.1 – A disc can become thermally unstable if its midplane temperature increases to $T_c \sim 10^4$ K; at this temperature, the opacity ($\kappa_R(T)$) rises quickly as a function of T due to hydrogen becoming ionised (***left***). In this situation, the thermal equilibrium function of the local vertical structure in the ($\log \Sigma$, $\log T_{\rm eff}$) plane has multiple solutions ("S-curve", ***centre***), with the top and bottom branches representing stable regions (quiescence, ***right***), while the middle branch – enclosed by the two inflection points – represents an unstable region (outburst). (Image: Armitage (2011))

density Σ and effective temperature $T_{\rm eff}$ implies that any increase/decrease in the viscous energy production rate is balanced by an increase/decrease in the energy loss rate), and the unstable middle one. Assuming that the annulus of the disc at r is located somewhere on the bottom branch of the "S-curve" shown in Figure 1.1, a small perturbation in $T_{\rm eff}$ would push the annulus r at a point above the "S-curve" where cooling is faster than heating, hence damping the perturbation (Hartmann and Kenyon, 1996). However, if there is a pile-up of material within the disc, its equilibrium point eventually reaches the lower of the two inflection points; here a small perturbation in $T_{\rm eff}$ causes the system to become thermally unstable, so instead it 'jumps' to the top branch of the curve, where the effective temperature is much higher. This, in turn, induces a large increase in the mass accretion rate for $\dot{M} \propto T_{\rm eff}^4$ (Hartmann and Kenyon, 1996) which gradually drains away the material from the disc, allowing the equilibrium point to travel down the top branch until the upper inflection point, where thermal instability sets in again, the disc 'jumping' back onto the stable, bottom branch. This process can be cyclic, assuming that the disc material is replenished (Hartmann and Kenyon, 1996).

Although there is little doubt on thermal instability playing an essential role

in triggering outbursts in FU Orionis variables, this can only be triggered in the very inner regions of the disc ($r \leq 0.1$ Astronomical Units (AU), Armitage et al. (2001)) and the models which assume it is the only responsible mechanism imply a very inefficient transport of angular momentum in the inner disc; indeed the most detailed of such models, by Bell and Lin (1994), yields an α-parameter value (a measure of the angular momentum transport due to turbulent motions; more details about it in Section 2.1.2) of only $\alpha \approx 10^{-3}$ when the midplane of the disc is ionised; this is however in disagreement with the observed disc lifetimes. Current models therefore also make use of the following effects in order to explain these aspects of FU Orionis outbursts:

- *Matter being accumulated in the inner disc*:
  - Interaction with a companion, disc-less star (Bonnell and Bastien, 1992; Pfalzner, 2008; Forgan and Rice, 2010).
  - Formation of clumps in the outer parts of the disc induced by gravitational instability, followed by its migration to the inner disc (Vorobyov and Basu, 2005; Vorobyov and Basu, 2010; Boley et al., 2010; Nayakshin, 2010)

  Although it is very much possible that external perturbations like the two described above can trigger an outburst, it is not clear whether their strength and recurrence can explain FU Orionis outbursts.

- *Timescale*: vertical stratification at $r \sim 1$ AU allows angular momentum to be transported by means of Magnetohydrodynamic (MHD) turbulence in the surface layers, where ionisation is present thanks to the action of stellar irradiation (X-ray or EUV/FUV) and cosmic rays. Given that the surface layers are characterised by a fixed accretion rate function (which depends on r) (Armitage et al., 2001), material from the gravitationally unstable outer regions builds up in the quiescent, dead zone. This in turn renders it unstable to gravitational instability, leading to a spur in angular momentum transport and an outburst. This continues until the build up of material is depleted and the disc cools back down from its previous hot, turbulent state; this allows for material to slowly build up in the dead zone again, matching the observed periodicity in FU Orionis outbursts (Gammie, 1999; Armitage et al., 2001).

Planet formation

Although only 1% of the initial disc mass is represented by solid particles (Williams and Cieza, 2011), their evolution is a key factor in understanding both the overall disc evolution and the eventual formation of planets within it, and it is illustrated in Figure 1.2.

As seen above, accretion onto the central star begins in the very early stages of the disc's lifetime, causing disc mass loss. Further mass loss occurs also through photoevaporation, mostly by means of energetic Far-Ultra Violet (FUV) photons (Gorti et al., 2009; Williams and Cieza, 2011), which limits the disc size (to a few hundred AU) by truncating its outer edge (Figure 1.2a).

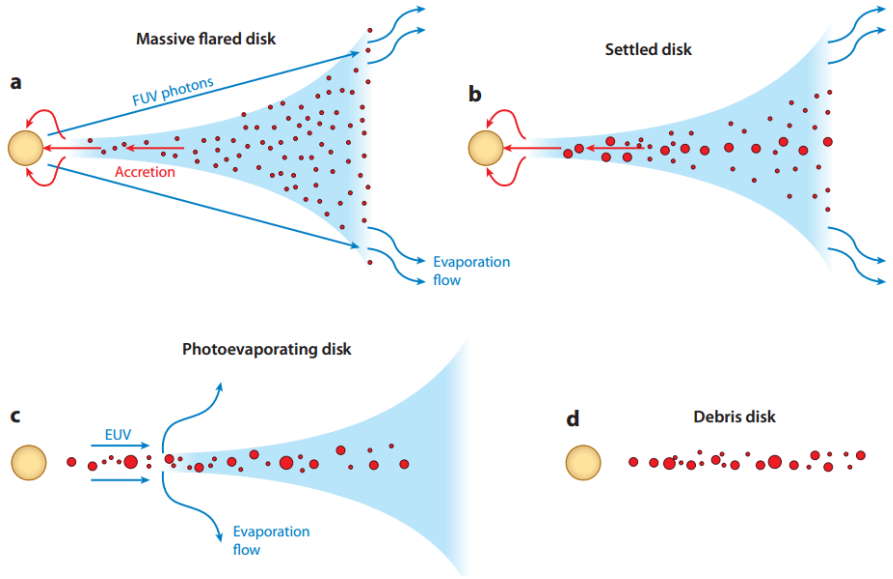


Figure 1.2 – Simplified evolution of a protoplanetary disc. ***(a:)*** Accretion onto the star and photoevaporation cause mass loss in the early stages of the disc, the latter delimiting the disc outer edge; ***(b:)*** Grains grow by collisions and decouple from the gas, settling towards the disc midplane, decreasing the scale height of the dust disc; ***(c:)*** As the accretion rate drops below the photoevaporation rate at small r, the inner region of the disc is depleted of material, forming a hole around the star, halting accretion. ***(d:)*** As photons can now hit upon the truncated inner edge of the disc, the photoevaporation rate increases further and the disc is quickly dissipated, leaving behind a debris disc where smaller grains are removed by radiative effects (Image: Williams and Cieza (2011))

Grains, which have a sub-micron size ($d \sim 0.1\mu$m), possess a large surface to mass ratio, meaning they are initially strongly coupled to the gas and swept along

with it. Inelastic collisions[2] with other grains ensure a drop in their surface to mass ratio, and a subsequent decoupling from the gas. No longer swept along by the turbulent motion of the gas, growing dust grains begin to settle towards the disc's midplane by being subject to a drag force. This settling of material increases the density of solid particles within the disc, resulting in a faster grain growth rate and a drop in the scale height of the dust (Figure 1.2b). In an idealised situation, this process would continue until all the planetesimals (or planetesimal embryos) would settle on the midplane, with only small grains lingering around the surfaces of the disc; however, due to the turbulent nature of PP discs, vertical mixing occurs (Dullemond and Dominik, 2005).

Photoevaporation, this time in the Extreme-UV range (EUV), becomes important at small r values when the accretion rate decreases; crucially, once this drops below the rate at which photoevaporation occurs, the inner disc can no longer be replenished with material (Alexander et al., 2006; Owen et al., 2010) and it is therefore depleted on a viscous timescale, creating a hole a few AU in size (Fig 1.2c). The formation of the aforementioned hole in the inner part of the disc allows photons to directly hit the truncated inner edge, further increasing the photoevaporation rate and effectively terminating/inhibiting accretion by preventing any material from filling the hole. Thereafter, the remaining gas quickly photoevaporates, leaving behind a debris disk dominated by radiative effects (Fig 1.2d); these radiative effects cause the destruction of micron-sized grains and a loss of angular momentum for larger ones (roughly sub-mm sized); this in turn prompts these larger grains to spiral in towards the central star (Poynting-Robertson effect) until they reach the dust sublimation radius, where they evaporate. This leaves behind a distribution of planets, planetesimals and planetesimal embryos.

1.2 Cataclysmic Variable systems

Cataclysmic Variables (CVs) are short-period systems of interacting binary stars consisting of a White Dwarf (WD) and a low-mass, near Main-Sequence star filling its Roche lobe, making this a type of semi-detached binary system.

Angular momentum loss – taking place through gravitational radiation and/or

[2]Dust dynamics are of course a much more complex phenomenon, as two colliding grains can merge with the total mass being retained by the new grain, bounce off each other with no mass transfer, or fragment with or without mass transfer, as explained in Testi et al. (2014)

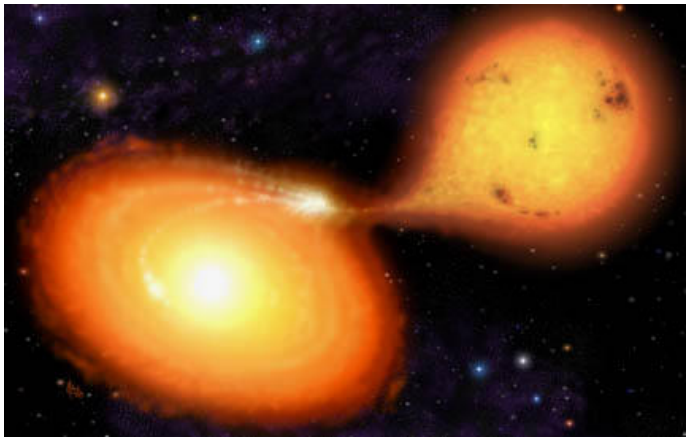


Figure 1.3 – Artistic representation of a CV system. The secondary star (on the right) loses mass due to the gravitational pull of the WD (on the left). This material spreads around the WD in an accretion disc, with new material causing a hot spot where it hits the disc (Image credit © Mark Garlick).

magnetic braking – causes the orbital separation between the two stars to decrease until eventually the near Main Sequence star (also called "the secondary") comes into contact with its Roche lobe (Kreidberg et al., 2010). When the outer layers of the secondary first overflow its Roche lobe, the star loses material through its inner Lagrangian point L_1; once the material leaves the secondary star's surface it can either leave the binary system altogether, return to the secondary or accrete onto the WD (Prendergast and Burbidge, 1968). In the latter case, if the WD (also called "the primary") is weakly magnetised, the material does not accrete directly onto the WD; instead, because it still possesses a substantial amount of angular momentum, it forms a ring around it at the circularisation radius; this is the radius at which the specific angular momentum is equal to that of matter leaving the secondary through the L_1 point. This ring is then subsequently broadened by viscosity into an accretion disc (Patterson, 1984), as shown in Figure 1.3.

Once the disc is in place, any further mass transferred to it from the secondary will impact the disc producing a 'hot spot', a shock front where energy is intensely dissipated (Robinson, 1976; Lin and Papaloizou, 1996). This is a prominent feature in the system's spectrum and its size and position have been heavily observed and researched (e.g. Warner and Peters (1972), Flannery (1974), Romanova et al. (2004), and Bąkowska and Olech (2014); the presence of the hot spot can also be deduced by the system's light curve; one example is the U Gem star, whose light curve shows a prominent 'hump' in both visible and UV light, indicating the high

temperature of its hot spot. A rapid, irregular flickering is also observed in the light curve, due to the unstable nature of the hot spot which manifests itself in an erratic luminosity (Robinson, 1976).
As the material begins to move inwardly through the disc, it releases gravitational energy heating up the disc to luminosities of $0.001-10L_{\odot}$ (Patterson, 1984), usually making the disc the most luminous part of a CV system (Townsley and Gansicke, 2008).

The accretion of material from the disc onto the primary can in some cases lead to outbursts in the system's lightcurve, which is the reason for the 'cataclysmic' part of the name for this type of system. Based on the intensity of these outbursts and their recurrence periods, CV systems have been subdivided into groups. The most important are as follows.

1.2.1 Classical novae

Classical novae take place in short period ($1 - 12$ hr) systems (José et al., 2012). These undergo short but very energetic outbursts, liberating energies greater than $10^{45}\mathrm{erg\,s^{-1}}$ (Trumper and Hasinger, 2008), inducing an increase in brightness of up to 6-19 magnitudes. These systems have been observed to undergo a single outburst.

Classical novae outbursts originate in a layer of accreted Hydrogen-rich material on the WD which becomes degenerate as a result of the compression caused by further accretion (José and Hernanz, 1998; Gallagher and Starrfield, 1978; Webbink et al., 1987; Parikh et al., 2014). As accretion continues, temperature and pressure increase at the bottom of the layer considered, eventually triggering a thermonuclear reaction converting Hydrogen into heavier elements by means of the CNO cycle (Trumper and Hasinger, 2008). The burning of Hydrogen atoms produces further heat, in turn increasing the nuclear burning rate causing a thermonuclear runaway.

The mechanism causing outbursts in classical novae is similar to that forming Type Ia supernovae. However, in classical novae the thermonuclear reaction occurs close to the surface of the WD, meaning the star is not destroyed during the event; because of this it is believed that classical novae are actually recurrent, with an estimated period of $\sim 10^4 - 10^5$ yrs (Prialnik, 1986; José et al., 2012).

1.2.2 Recurrent novae

Recurrent novae are a small subgroup of CVs (~ 10 confirmed targets) which show multiple outburst events resembling those seen in classical novae, though of smaller intensity. Unlike classical novae, recurrent novae have been observed to undergo several outbursts, with recurrence periods of $10 - 100$ yrs, although the upper end of the range is certainly affected by observational constraints. Figure 1.4a shows a lightcurve for a recurrent nova, with a recurrence period of roughly 10 years.

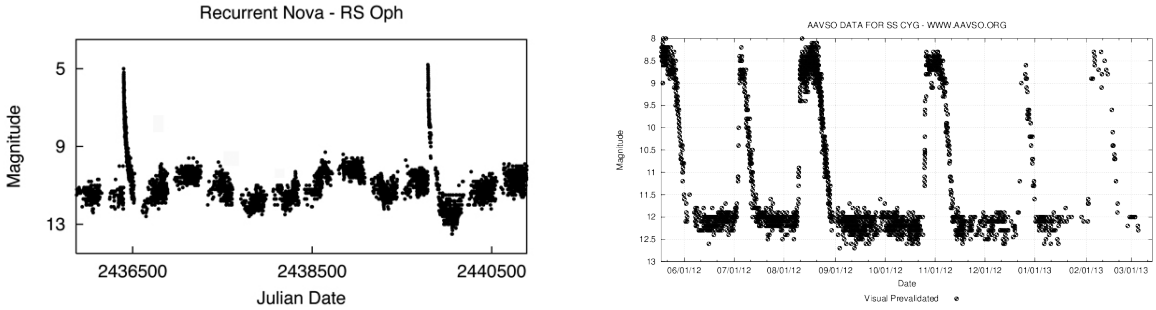
Although the mechanism triggering the outbursts is, just like in classical novae, thermonuclear runaway at the bottom of an accreted H-rich shell of material, the smaller amount of energy released might have a profound effect on the fate of such systems. Observations support theoretical predictions by detecting supersoft X-ray emission $\sim 10 - 60$ days after the optical outbursts, implying nuclear burning carries on for this period (Yungelson et al., 1996; Kahabka et al., 1999). This means that, unlike classical novae where the WD mass decreases after an outbursts, in recurring novae only a fraction of the accreted mass is ejected following thermonuclear runaway. With the primary's mass increasing with time, recurring novae have become a focus of interest as progenitors of Type Ia Supernovae (Kahabka et al., 1999; Bode, 2011).

1.2.3 Dwarf novae

Dwarf novae are the most populous and researched type of CV; these systems show smaller outbursts, with the system's brightness increasing by a factor of roughly 100. The outbursts are however much more frequent compared to the other categories of novae discussed above; in fact, their outbursts are usually separated by only $\sim 50 - 100$ days (Hellier, 2001; Kreidberg et al., 2010), as shown in Figure 1.4b.

Unlike classical and recurrent novae, where outbursts are due to material being piled up on the WD, dwarf novae outbursts are due to disc luminosity. This can be inferred from the lightcurve and spectrum of the system during the outburst, which gradually transition from showing significant contributions from the WD and the hot spot to being dominated by the disc component (Hellier, 2001).

Two major models were developed to attempt explain outbursts in dwarf novae; one suggests that accretion onto the WD is caused by increased mass transfer from the secondary to the disc, causing an increase in the disc's temperature and



(a) **(b)**

Figure 1.4 – ***(a)*** Light curve of recurrent nova RS Oph showing two outbursts of about 6-7 magnitudes, roughly ten years apart (Credit: AAVSO). ***(b)*** Light curve for the dwarf nova SS Cyg showing semi-regular, small outbursts, lasting $1-2$ weeks, with quiescent phase interludes of $4-10$ weeks. (Credit: AAVSO).

luminosity. The opposing model is that proposed by Osaki (1974), where the escalation in the disc's luminosity is due a thermal instability within the disc; the model states that the reason for the disc instability is a pile-up of material due to the rate of mass transfer from the secondary to the disc being larger than the rate of transfer through the disc itself (Hellier, 2001). The ensuing instability boosts the viscosity, and therefore the disc's angular momentum transport. This causes the disc to spread both outwards and inwards, leading to accretion onto the WD and an increased luminosity in the disc (Hellier, 2001; Kreidberg et al., 2010). Following the outburst, during which the disc is drained of matter, the disc returns to a quiescent, low-viscosity state during which its matter is continuously replenished by mass transfer from the secondary, readying the disc towards the next outburst.

Observational evidence has since strongly favoured the disc instability model (Buat-Ménard et al., 2001; Kreidberg et al., 2010). One of the main issues with the opposing model is that, were the outburst triggered by an enhanced mass transfer from the secondary, the hot spot's luminosity should also increase during outbursts; observations however show that it remains roughly constant (Hellier, 2001). Furthermore, the model by Osaki (1974) correctly predicts the spreading and contracting of the disc during and after an outburst, respectively, the latter being due to matter transferred from the secondary having a specific angular momentum (corresponding to the circularisation radius) lower than that possessed by material in the outer part of the disc (Livio and Verbunt, 1988; Hellier, 2001). In the enhanced mass transfer model the disc will shrink at the start of an outburst for

the same reason, eventually spreading again during quiescence, hence contradicting the observations (Livio and Verbunt, 1988; Lewin et al., 1997; Hellier, 2001).

1.3 Active Galactic Nuclei

Some galaxies possess a central region whose luminosity can outshine that of all stars in the rest of the galaxy, with an emission spectrum bright across all wavelengths. These central regions, called Active Galactic Nuclei (AGN), host a Supermassive Black Hole (SMBH); the radiation emitted by the AGN (which have different nomenclatures depending on their ratio of radio-to-optical flux density, optical emission spectrum properties and, importantly, their orientation to the line of sight) is thought to be caused by accretion driven by the gravitational potential of the SMBH. The aforementioned accretion onto the SMBH can have significant consequences on the properties of its host galaxy as the SMBH feeds back energy in the form of radiative and kinetic energy. The presence of this feedback was first realised thanks to what is known as the $M_{\mathrm{BH}} - \sigma$ relation, showing a remarkably strong correlation between the mass of the central SMBH and the bulge stellar velocity dispersion σ (Ferrarese and Merritt, 2000; Gebhardt et al., 2000). A similarly powerful correlation also exists between the SMBH's mass and the mass of the galactic bulge (Marconi and Hunt, 2003; Haring and Rix, 2004).

Models of AGN accretion seem to agree on the existence of two distinct types of accretion: cold mode and hot mode (Hardcastle et al., 2007; Zinn et al., 2013). Gas from the galaxy merger fuels the efficient cold mode accretion, with its X-ray emission spectrum being dominated by the disc component; star formation is plentiful during this stage. As the black hole accretes matter and eventually reaches its critical mass (given by the $M_{\mathrm{BH}} - \sigma$ relation), its feedback onto the surrounding regions clears gas and material out of the galaxy by heating it up and disrupting it. This causes accretion onto the SMBH to run out of fuel and stop and star formation within the galaxy to halt, matching the observed quenching in star formation rate. The AGN's disruptive action on the star forming region is called negative feedback. With the cold gas from the galaxy merger having been accreted onto the SMBH, the latter can only accrete hot gas from the system's halo – if accretion takes place at all – in what is called hot mode accretion (also referred to as "radio mode" accretion as a large part of radio AGN are observed to be accreting through this mode, or Bondi accretion (Zinn et al., 2013)).

The feedback caused by SMBH accretion onto the star forming regions of the galaxy can also however be positive. Observations of the BR1202-0725 quasar made with the NICMOS3 camera from the Hubble Space Telescope show a correlation between the most active star formation regions of the galaxy with the axis of the radio jet launched by the AGN (Klamer et al., 2004). This is believed to be caused by the compression induced by the jet during its propagation through the gas, generated by either shocks or turbulent motions; this in turn caused the gravitational collapse of marginally stable regions into stars.

1.4 X-ray Binaries

Binary systems in which one of the two stars goes supernova, turning into either a neutron star (NS) or a black hole (BH), are called X-ray Binaries (XB) as the gravitational potential of either NS or BH – which is stronger than that of a WD in a CV – causes energy emission up to the X-ray range. In the case where the accretor is a NS, the magnitude of its magnetic field greatly influences the accretion dynamics: if the magnetic field is relatively weak ($B < 10^6$ T) the disc may be truncated and the emission spectrum is dominated by the inner disc and the boundary layer. If, on the other hand, the NS possesses a strong magnetic field ($B \sim 10^8$ T) or rotates rapidly, any accretion flow within a few hundred times of the NS radius would be disrupted, with the accreting material being channelled by magnetic field lines and accreted onto the magnetic poles of the NS, which will emit X-rays (Pringle and Rees, 1972; Ostriker and Davidson, 1973; Lamb et al., 1973; Lewin et al., 1997). If the rotation and magnetic axes happen to be misaligned and the X-ray emission sweeps through the line of sight, this would be perceived as a pulsation (Meszaros et al., 1980; Wang and Welter, 1981). If the accreting object is a BH, the disc can be the main contributor to the X-ray emission if the mass flux is within a defined range; if the mass flux is too small or too large, accretion onto the BH would occur in bursts followed by the BH expelling material by means of flares (Pringle and Rees, 1972).

The mass of the companion star M_* determines whether the binary system is categorised as a Low-Mass X-ray Binary (LMXB, $M_* \lesssim 1M_\odot$) or High-Mass X-ray Binary (HMXB, $M_* > 10M_\odot$).

High-Mass X-ray Binaries (HMXBs)

High-mass X-ray binaries are initially composed of two massive stars, the more massive of which moves more quickly through its evolutionary stages. Once this more massive star (the 'primary') evolves into a giant, it sheds its outer layers into the interstellar medium and a part of this ejected material is accreted onto the less massive star (the 'secondary'), shifting the mass distribution within the binary system. As the star goes supernova creating a NS or a BH, the asymmetry of this collapsive phenomenon means that the dying star receives a velocity kick that – depending on its strength – can significantly affect the orbital characteristics of the system or even overcome the two stars' gravitational pull, unbinding the binary (Coe, 2005). Mass transfer in this case does not occur through Roche lobe overflow as in the case for the CVs, but through the capture of the stellar wind emitted by the secondary (having a rate of $10^{-6} - 10^{-10} M_{\odot}\,\mathrm{yr}^{-1}$), which is driven by its radiation pressure (Lewin et al., 1997); it is the conversion of the accreting wind material's potential energy into radiation that produces the observed hard X-ray signature.

Low-Mass X-ray Binaries (LMXBs)

Low-mass X-ray binaries consist of NS or BH and a companion low-mass star, such as a late-type star, a WD or a brown dwarf, locked in a very tight configuration. Although it is still not clear how LMXBs form, the current view is that their formation requires the two stars to be in a common envelope set up before evolution turns the more massive one into a BH or NS (Naoz et al., 2016). The binary then retains its close configuration so that the low-mass companion is able to fill its Roche lobe and shed material through its inner Lagrangian point, forming a disc around the compact object.

1.5 Gamma-ray bursts (GRBs)

Gamma-ray bursts are short, extremely energetic explosions consisting of collimated beams of low-energy γ rays, followed by long-lasting afterglows visible in X-ray, optical, IR and radio wavelengths. These bursts are usually divided into two categories: short bursts (if their duration is $\Delta t < 2\,\mathrm{s}$) or long bursts ($\Delta t > 2\,\mathrm{s}$) (Piran, 2004).

By studying the properties of GRBs – such as variability, energy outputs, collimation angles and burst durations – it is possible to infer that they are produced by rapid accretion onto a compact object, probably a black hole, surrounded by a massive disc ($M_{\rm disc} \sim 0.1 M_{\odot}$). What however remains in doubt is the mechanism that triggers such an extreme scenario of accretion onto a compact object; in addition, the fact that some GRBs exhibit short bursts while others show long ones has been interpreted by some as it being proof of there being two different mechanisms causing GRBs (Narayan et al., 2001)

1.6 Generality of the work

As mentioned previously, the work carried out in this thesis considers a generic Keplerian accretion disc, making few assumptions that might constrain its nature or that of the hosting system. One of such few constraints is the consideration of the disc's own self-gravity. While of course this represents a generalisation in itself, considering systems where self-gravity has an important contribution restricts these analyses to AGN discs and early stage protoplanetary discs; self-gravity is however not considered at all times. The Keplerian rotation profile of the disc in question is of course a mild constraint in itself, ruling out more extreme systems such as some galactic discs. However zonal flows, which are the central phenomenon to the work and common throughout the different analyses carried out, can develop in any type of disc.

Chapter 2

Dynamics of accretion discs

The evolution of accretion discs is centred around the transport of angular momentum within it. This occurs due to a process analogous to a viscous torque, which however has a macroscopic origin (i.e. it operates on scales comparable to the scale height of the disc, rather than on scales of order of the particles' mean free path); this causes angular momentum to be transported in the radial direction, perpendicular to the flow direction. In order for accretion to occur, angular momentum must be transported outwards, meaning that the matter located in the inner regions of the disc will lose angular momentum and consequently move to an orbit closer to the central object as the (specific) angular momentum j is

$$j(r) = r^2\Omega = \sqrt{GM_c r}, \tag{2.1}$$

where G is the gravitational constant, r is the distance from the central object possessing mass M_c and Ω the angular velocity of a Keplerian disc

$$\Omega(r) = \sqrt{\frac{GM_c}{r^3}}. \tag{2.2}$$

It is the continual repetition of this process that brings material to spiral inwards onto the central object, causing accretion to take place. The outward transport of angular momentum has therefore the consequence of transporting mass inward. Moreover, as the particles spiral inwards to lower (more negative) energy orbits, they lose gravitational energy; half of the lost amount of gravitational energy is converted into kinetic energy – as orbital angular velocity increases with decreasing distance from the central object – whilst the other half is converted into heat. Most

of this heat then escapes the disc, cooling it down.

Such transport of angular momentum can be induced by several physical mechanisms, as explained in more detail in Section 2.2, such as turbulence arising from disc instabilities of magnetohydrodynamical and hydrodynamical nature as well as other processes such as self-gravity, if the disc is massive enough.

2.1 The importance of dissipative effects

2.1.1 The example of a ring of material

The action of viscosity upon fluid particles therefore plays a crucial role in the dynamics of an accretion disc. To understand how viscosity induces the transport of mass mentioned above, let us consider a circular disc (Lynden-Bell and Pringle, 1974; Pringle, 1981), whose evolution is dictated by mass and angular momentum conservation equations, their height-integrated versions being

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}\left(r\Sigma v_r\right) = 0, \tag{2.3}$$

$$\frac{\partial}{\partial t}\left(r^2\Sigma\Omega\right) + \frac{1}{r}\frac{\partial}{\partial r}\left(r^3\Sigma\Omega v_r\right) = \frac{1}{r}\frac{\partial}{\partial r}\left(r^2 W_{r\phi}\right), \tag{2.4}$$

where Σ is the surface density of the disc, v_r the velocity in the radial direction and $W_{r\phi}$ the (dominant) $r\phi$ component of the viscous stress tensor

$$W_{r\phi} = \nu\Sigma\frac{\mathrm{d}\Omega}{\mathrm{d}\ln r}. \tag{2.5}$$

Equations 2.3 and 2.4 can then be combined to give an equation for the viscous evolution of the surface density for a Keplerian disc

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{r}\frac{\partial}{\partial r}\left[r^{1/2}\frac{\partial}{\partial r}\left(\nu r^{1/2}\Sigma\right)\right]. \tag{2.6}$$

This is the non-linear (as the viscosity ν is a function of Σ) diffusive equation of the disc, whose radial velocity

$$v_r = -\frac{3}{\Sigma r^{1/2}}\frac{\partial}{\partial r}\left(\nu\Sigma r^{1/2}\right) \sim \frac{\nu}{r} \tag{2.7}$$

is proportional to the viscosity ν – where the latter can potentially be a rather complicated function of several local variables, such as temperature, radius and surface density – and the characteristic viscous timescale is

$$\tau_{\rm visc} \approx \frac{r^2}{\nu} \sim \frac{r}{v_r}. \tag{2.8}$$

In discs the viscous timescale is generally much longer than the dynamical timescale $\tau_{\rm dyn} = \Omega^{-1}$, therefore implying that the rotational velocity of the disc is much larger than the radial velocity.

In order to better illustrate the effects of viscosity upon the evolution of the disc, we consider the special case where $\nu = \text{const.}$; this allows us to solve Equation 2.6 analytically by means of Green's function. By assuming that at $t = 0$ the gas lies solely in a thin ring of mass m located at $r = r_0$, the initial surface density of the gas is defined by

$$\Sigma(r, t = 0) = \frac{m}{2\pi r_0} \delta\left(r - r_0\right), \tag{2.9}$$

where $\delta\left(r - r_0\right)$ is a Dirac delta function centred on r_0. Allowing for free expansion of the gas as $r \to \infty$, the solution is given by (Lynden-Bell and Pringle, 1974)

$$\Sigma(r, t) = \frac{m}{\pi r_0^2} \frac{1}{\tau} x^{-1/4} \exp\left[-\frac{(1 + x^2)}{\tau}\right] I_{1/4}\left(\frac{2x}{\tau}\right), \tag{2.10}$$

where the surface density is expressed in terms of the dimensionless parameters $\tau \equiv 12\nu r_0^{-2} t$ and $x \equiv r/r_0$, and $I_{1/4}$ is a modified Bessel function of first kind. The solution to the equation is shown in Figure 2.1 for a range of dimensionless times τ.

As τ increases the ring is diffusively spread to both inner and outer orbits, with mass eventually flowing to $r = 0$ and accreting onto the central object. The spreading of the ring, which occurs in order to conserve angular momentum, appears asymmetric, with most matter losing energy and angular momentum and ending up in the inner region of the disc (or accreting), while a small fraction of the mass carries the bulk of the angular momentum towards $r = \infty$. Regardless of this redistribution of the system's angular momentum, its total value is conserved. Further analysis of the solution shows that the largest radius at which matter is moving inward increases with time according to $r_{\rm max} = t\nu/r_0$ (Pringle, 1981; Frank et al., 2002; Lodato, 2007). As a result it is possible to predict that as $t \to \infty$

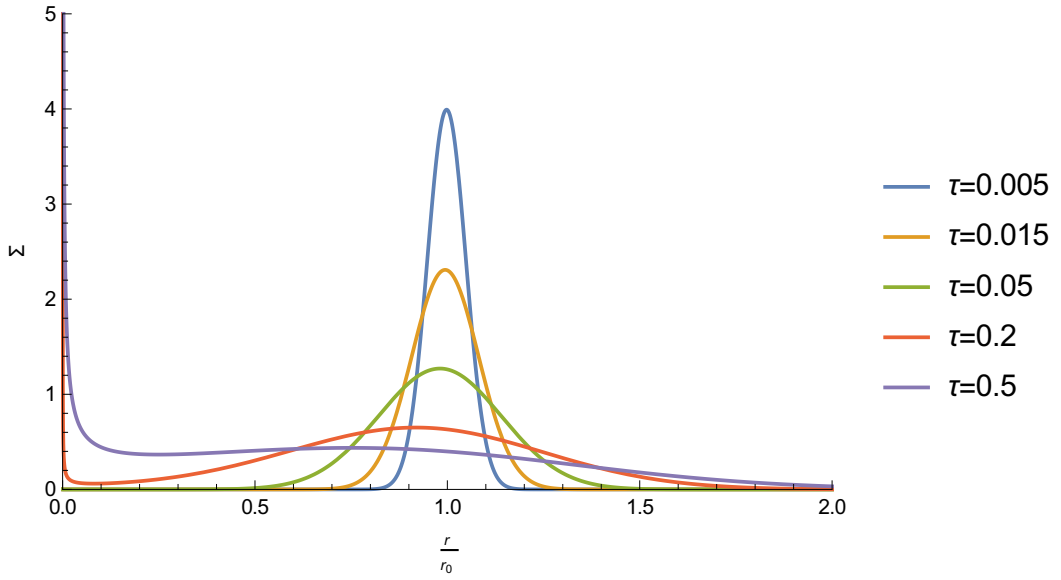


Figure 2.1 – Viscous evolution of a ring of gas with constant viscosity ν initially located at a radial distance r_0 for different times τ. The action of viscosity acts to spread the ring in both directions (asymmetrically preferring the inward direction) thereby leading to accretion onto the central object.

almost all of the disc's original mass would have lost its angular momentum and would have therefore accreted onto the central object, while a negligible amount of material would have moved to $r \to \infty$, carrying all the angular momentum.

2.1.2 The α-disc model

The above analysis suggests that a viscosity-like diffusive process which transports angular momentum outwards induces an inward transport of mass, therefore causing accretion onto the central object on a viscous timescale $\tau_{\rm visc}$. The physical nature of the viscosity responsible for angular momentum transport is however still source of debate. A disc's molecular viscosity $\nu_{\rm mol}$, which is given by

$$\nu_{\rm mol} \sim \lambda_{\rm mol} v_{\rm mol}, \tag{2.11}$$

where $\lambda_{\rm mol}$ is the collisional free path of the molecules and $v_{\rm mol}$ their (random) velocity, is in fact too small to explain the observed accretion rates and disc lifetimes. Indeed, in the case of a proto-planetary disc, typical values of the number density n and molecular cross-section $\sigma_{\rm coll}$ are $n \sim 10^{13}\,{\rm cm}^{-3}$ and $\sigma_{\rm coll} \sim 10^{-16}\,{\rm cm}^2$,

yielding a typical mean free path $\lambda_{\rm mol} = 1/(n\sigma_{\rm coll}) \sim 10^3\,{\rm cm}$. The ratio of viscous to dynamical timescales is

$$\frac{\tau_{\rm visc}}{\tau_{\rm dyn}} = \frac{r^2\Omega}{\nu_{\rm mol}} = \frac{r^2}{H\lambda_{\rm mol}} = {\rm Re}, \tag{2.12}$$

where $H \simeq c_{\rm iso}/\Omega$ is the scale height of the disc, with $c_{\rm iso} \sim v_{\rm mol}$ being the isothermal sound speed of the gas, and

$${\rm Re} \equiv r^2\Omega/\nu \tag{2.13}$$

is the Reynolds number of the flow for a generic kinematic viscosity. If the disc is assumed to have a radial extent $r \sim 100$ AU and an aspect ratio $H/r \sim 0.1$, the ratio of viscous to dynamical timescales is found to be

$$\tau_{\rm visc}/\tau_{\rm dyn} = {\rm Re} \sim 10^{13};$$

as the dynamical timescale $\tau_{\rm dyn} \sim$ few years for proto-planetary discs, this means that the viscous timescale – the time during which all of the disc matter accretes onto the central object – is of the order $\tau_{\rm visc} \sim 10^{13}$ yrs; this is however much longer than the observational lifetime of proto-planetary discs, which are found to disperse in their entirety within $\tau_{\rm lifetime} \sim 10^6 - 10^7$ yrs. It seems therefore likely that, if accretion plays a significant part in limiting the discs' lifetime, the effective viscosity responsible for the angular momentum transport must come from some other physical process, which would provide a larger value of ν.

The extremely high Reynolds numbers (${\rm Re} \sim 10^{13}$) associated with these discs does hint at the possibility of turbulence arising in their flows, as laboratory experiments have found that hydrodynamic flows show a transition to a turbulent state if the Reynolds number of the flow is sufficiently large (but smaller than the value required to induce a linear instability, hence why the transition to turbulence is called *sub-critical*) (e.g. Taylor, 1936; Dauchot and Daviaud, 1994; Longaretti, 2002). Turbulence does however appear to be suppressed in planar Couette flows as soon as a Coriolis force is introduced (Balbus et al., 1996; Hawley et al., 1999), although this disagrees with some results obtained using Taylor-Couette flows (Wendt, 1933; Taylor, 1936; Richard, 2001; Dubrulle et al., 2005). Nevertheless, the Reynolds numbers attained in accretion discs are several orders of magnitude larger than those achieved in laboratory experiments, so it seems pos-

sible that turbulence could arise in discs. Should that be the case, the angular momentum transport would occur as a consequence of turbulent mixing of fluid elements.

By considering a background Keplerian flow with a turbulent perturbation, it is realistic to assume that the averaged turbulent stress is of a similar form of the viscous stress $W_{r\phi}$ defined in Equation 2.5 and that the resulting effective turbulent viscosity can be generally estimated by

$$\nu_{\rm turb} \sim \lambda_{\rm turb} u_{\rm turb}, \tag{2.14}$$

where $\lambda_{\rm turb}$ and $u_{\rm turb}$ are the characteristic length-scale and velocity of the turbulent state, mirroring the expression for the molecular viscosity given in Equation 2.11. The issue however is that turbulence is a very complex, still poorly-understood phenomenon, so it is not known what values its characteristic length-scale and velocity would take in this context.

The seminal work by Shakura and Sunyaev (1973) made use of simple physical arguments to bypass this problem. Firstly they limited the physical size of the turbulent eddies to the disc's scale height H (i.e., $\lambda_{\rm turb} \lesssim H$); this is an important assumption as otherwise turbulent transport would have a global nature and it could not be described by means of a viscosity coefficient. The second assumption made is that the characteristic speed of turbulent motions cannot be very supersonic (i.e., $u_{\rm turb} \lesssim c_s$); were that the case, strong shocks would ensue and damp out the turbulent motions to subsonic speeds. By applying these two assumptions, Shakura and Sunyaev (1973) parameterised the turbulent viscosity in the following way

$$\nu_{\rm turb} = \alpha c_s H, \tag{2.15}$$

where α is a dimensionless free parameter whose value is less than unity owing to the assumptions mentioned above. The above expression, which is valid for thin discs ($H \ll r$) is often referred to as the 'α-prescription' of turbulent viscosity as the parameter α encloses all the uncertainties regarding the properties of turbulence-driven viscosity. The advantage of this scaling law is to replace the unknown turbulent viscosity $\nu_{\rm turb}$ with the measurable dimensionless α parameter; this varies more slowly than $\nu_{\rm turb}$ with disc properties (e.g. radius, temperature) and encloses all the uncertainties regarding accretion disc turbulence. Further-

more, the α parameterisation stands and can be applied regardless of the source of viscosity. What's more, it allows a direct link to observations of discs as a specific value of α (and therefore a specific rate of angular momentum transport) allows the estimation of the disc's lifetime, which can be directly compared to the observed values. Although analytical models usually employ a constant value for the α parameter, the picture is of course somewhat more complicated in numerical simulations, with α showing both temporal and spatial variations (e.g. Papaloizou and Nelson, 2003).

Although this very simple expression in some way hides the shortcomings in the research of turbulent dynamics, it is crucial to better understand how turbulence operates, as it gives us a better idea of how the value of α varies within the disc and whether assuming a constant α is a realistic simplification of the problem. One way to deepen our knowledge of turbulence is to pinpoint how it arises in the first place. What we do know about turbulent motions in discs is that they tend to extract energy from the largest scales (i.e. $\sim H$) and cascade it down to progressively smaller scales, until it is eventually dissipated. The initial energy extraction is possible thanks to the action of disc instabilities. Such mechanisms can either be driven by the action of a magnetic field, like the Magnetorotational Instability (MRI) – in which case they are labelled Magnetohydrodynamic (MHD) mechanisms – or be non-magnetic in nature; the latter type of mechanisms include hydrodynamical instabilities, Rossby wave instability, convection and gravitational instability amongst others. It is important to be aware that not all these instabilities produce a turbulent state in its classical definition, as some of the aforementioned mechanisms give rise to more orderly non-linear phenomena such as convective rolls and coherent vortices. Nevertheless, these phenomena can still be modelled according to Equation 2.14 as giving rise to a general turbulent viscosity, where in this case $\lambda_{\rm turb}$ and $u_{\rm turb}$ would take the values of that particular phenomenon's characteristic lengthscale and velocity.

2.2 Angular momentum transport mechanisms

2.2.1 Magnetorotational Instability (MRI)

The MRI is a linear instability that arises in ionised regions of discs, which are threaded by a magnetic field, due to their differential rotation; it operates on a

dynamical timescale, $\tau_{\rm MRI} \sim 1/\Omega$. The MRI was initially discovered in the late '50s by Velikhov (1959) following his work on the stability of Taylor-Couette flows (a flow confined in the annulus separating two concentric, rotating cylinders) threaded by a vertical magnetic field; MRI dynamics were also pursued and examined in more detail by Chandrasekhar (1960). The topic was occasionally revisited in the following decades, with applications in stellar differential rotation and geodynamos, but it was not until the '90s that the pioneering work by Balbus and Hawley (1991) applied the idea of MRI to differentially rotating accretion discs, provided they were threaded by a magnetic field and were sufficiently ionised.

The MRI is the consequence of two basic properties: the frozen flux approximation and the disc's differential rotation. The former is the property of fluids having an ideal electric conductivity in which magnetic field lines are frozen into the fluid itself. This characteristic – which is described by Alfvén's theorem – means that if two fluid elements which are both threaded by the same field line were to undergo a relative displacement, the magnetic field line itself would be forced to distort to follow the displacement. The two fluid elements could then be likened to being linked by a spring. Usually, the restoring force from the spring would act to neutralise the displacement between the fluid particles; this however is not the case if the system exhibits a differential rotation, as schematically represented in Figure 2.2.

If two fluid elements have some initial radial separation (**a**), the differential nature of the disc's rotation increases the displacement (**b**) as the two fluid elements would possess different angular velocities (the inner element possessing a greater angular velocity although the outer element has a greater amount of angular momentum). The tension of the frozen-in magnetic field line allows angular momentum to be exchanged between the two elements by exerting a retarding torque on the element ahead (blue) and a positive torque on the element trailing behind (red); this causes angular momentum to be exchanged in the outward direction, resulting in the inner parcel losing angular momentum and dropping to an orbit closer to the central object, while the outer parcel is pushed to an outer orbit (**c**). This causes the separation between the two fluid elements to increase, with the process repeating itself in a runaway fashion as an increasing separation leads to progressively larger tension in the spring and, therefore, increasing exchange of angular momentum between the elements.

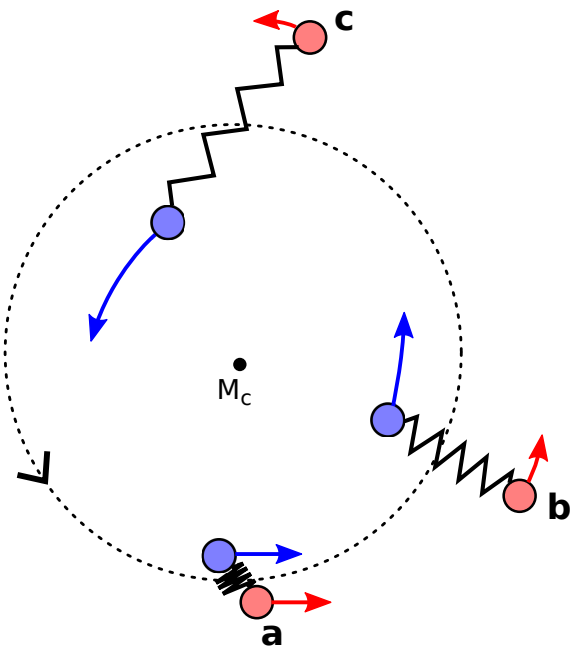


Figure 2.2 – Schematic representation of how the MRI operates: the initial displacement between two fluid elements is increased by the differential nature of the disc's rotation and by the outward transport of angular momentum taking place through the frozen-in magnetic field line.

Linearised MHD equations can be used to derive the following local dispersion relation for small axisymmetric disturbances (assuming a solution of the form $\mathrm{e}^{-\mathrm{i}\omega t}$)

$$\omega^4 - \omega^2 \left[\kappa^2 + 2\left(\boldsymbol{k}\cdot\boldsymbol{v_A}\right)^2\right] + \left(\boldsymbol{k}\cdot\boldsymbol{v_A}\right)^2 \left[\left(\boldsymbol{k}\cdot\boldsymbol{v_A}\right)^2 + \frac{\mathrm{d}\Omega^2}{\mathrm{d}\ln r}\right] = 0, \tag{2.16}$$

where ω is the frequency of the disturbance and $\boldsymbol{k}$ its wavenumber, $\boldsymbol{v_A} = \boldsymbol{B}/\sqrt{4\pi\rho}$ is the Alfvén velocity, $\boldsymbol{B}$ the magnetic field threading the disc and κ the epicyclic frequency given by

$$\kappa^2 = \frac{2\Omega}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\Omega\right). \tag{2.17}$$

By setting $\omega^2 = 0$ it is possible to obtain the criterion for MRI instability (for which $\omega^2 < 0$)

$$k^2 v_A^2 + \frac{\mathrm{d}\Omega^2}{\mathrm{d}\ln r} < 0. \tag{2.18}$$

This clearly shows that MRI cannot operate in a system with rigid body rotation, as

a necessary condition for the instability is for the angular velocity to be a decreasing function of radius, such that

$$\frac{\mathrm{d}\Omega^2}{\mathrm{d}\ln r} < 0, \tag{2.19}$$

which is true for typical accretion discs and Keplerian rotation profile discs. It is interesting to notice that this expression, which also represents the simplified instability criterion in the case of $k \to 0$ or in the weak field limit, is independent of the magnetic field $\boldsymbol{B}$; this implies that there is no threshold $\boldsymbol{B}$ strength for MRI to operate.

The maximum growth rate for the instability is found to be

$$|\omega_{\max}| = \frac{1}{2}\left|\frac{\mathrm{d}\Omega}{\mathrm{d}\ln r}\right|, \tag{2.20}$$

which is related to the angular velocity profile and is again independent of the magnetic field strength. As mentioned above, this indicates that the MRI can theoretically operate for an arbitrarily small $\boldsymbol{B}$; what the strength of the magnetic field influences is the perturbation wavenumber to which the maximum growth rate corresponds, which is given by

$$(\boldsymbol{k}_{\max} \cdot \boldsymbol{v_A})^2 = -\left(\frac{1}{4} + \frac{\kappa^2}{16\Omega^2}\right)\frac{\mathrm{d}\Omega^2}{\mathrm{d}\ln r}; \tag{2.21}$$

what this indicates is that the weaker the magnetic field, the smaller the wavelength triggering the largest growth rate. This means that the MRI has a local nature for weak magnetic field configurations.

Since the 're-discovery' of MRI by Balbus and Hawley (1991), its onset and evolution have been heavily researched in various different configurations in both local (Brandenburg et al., 1995; Stone et al., 1996; Sano and Stone, 2002; Fromang and Papaloizou, 2007; Bodo et al., 2008; Longaretti and Lesur, 2010, and many more) and global models (e.g. Armitage, 1998; Hawley, 2001; Papaloizou and Nelson, 2003). The simulations consistently verified that in the non-linear regime MRI leads to a turbulent state, which transports angular momentum towards the outward regions of the disc, mostly by means of Maxwell stresses. The simulations found that the resulting value of the α parameter was dependent on various factors – such as (explicit or numerical) viscosity, numerical resolution and vertical strat-

ification – and was typically in the $\alpha \approx 10^{-3} - 10^{-1}$ range; this is roughly in line with the values extrapolated from observations.

It is thanks to this consistent array of results that MRI is today considered to be the prime candidate for turbulence in accretion discs.

2.2.2 Non-magnetic mechanisms

As mentioned above, the triggering of the MRI requires a sufficient (but very low) level of ionisation, which can usually be provided by external sources such as cosmic rays and stellar X-rays (Stoll and Kley, 2014). However, there do exist discs – or regions of discs – where the level of ionisation is insufficient to trigger any magnetically-driven instability, as the magnetic field and the gas are poorly coupled. One such region is the so called 'dead zone', which in the vertically stratified disc models is centred around the disc's mid-plane and sandwiched by a surface layer on each side, which prevents ionisation from external sources from reaching it. The ionisation level in the dead zone region is therefore too low for MRI to operate, and we therefore require an alternative source of turbulence to explain the accretion rates.

Below is a short collection of non-magnetic instabilities which have been linked to the development of turbulent motions within the disc.

Hydrodynamical instabilities

The first alternative to MHD instabilities in regions which are not sufficiently ionised is a hydrodynamic instability. Theoretically speaking, there exist flows which exhibit linear instabilities when Re exceeds a critical value $\mathrm{Re}_c^{\mathrm{lin}}$ (e.g. plane Poiseuille flow), while others are linearly stable for all values of the Reynolds number (e.g. plane Couette flow), meaning their critical Reynolds number $\mathrm{Re}_c^{\mathrm{lin}} = \infty$. The Keplerian rotation profile, which represents a good approximation of real discs' profiles, presents a radially-increasing angular momentum profile, making it stable to linear instabilities according to Rayleigh's criterion (Rayleigh, 1880). Laboratory experiments show however a slightly different picture, with turbulence setting in for Reynolds number threshold values $\mathrm{Re}_c^{\mathrm{obs}}$ which are smaller than those predicted by the hydrodynamical models. Taking this into consideration, it appears that some effect must be triggering a *sub-critical* transition to turbulence for $\mathrm{Re}_c^{\mathrm{obs}} < \mathrm{Re} < \mathrm{Re}_c^{\mathrm{lin}}$ (Mukhopadhyay et al., 2005).

A possible mechanism responsible for the sub-critical transition to turbulence is the so-called *by-pass* transition, which originated in the hydrodynamical community in the 1990s (e.g. Trefethen et al., 1993; Gebhardt and Grossmann, 1994; Baggett et al., 1995; Waleffe, 1997; Grossmann, 2000; Schmid, 2000; Reshotko, 2001) and was later picked up by the astrophysics field as a possible mean to explain non-magnetic disc turbulence (e.g. Longaretti, 2002; Chagelishvili et al., 2003; Yecko, 2004; Mukhopadhyay et al., 2005; Afshordi et al., 2005; Lesur and Longaretti, 2005; Rincon et al., 2007). This transition to a turbulent state is believed to occur thanks to the non-normality of linear operators in a shear flow; as the angle between two modes' eigenvectors goes as

$$\cos\phi = \mathrm{Re}/\sqrt{\mathrm{Re}^2 + |\Delta\lambda_{\mathrm{ev}}|^2},$$

where $\Delta\lambda_{\mathrm{ev}}$ is the difference between the two eigenvalues, the modes become more and more parallel as the Reynolds number is increased. In high-Re regime, and if the modes are roughly orthogonal to a perturbation in the flow, a combination of linear modes can exhibit a transient energy growth; this is because, although all the linear modes are individually decaying, they do so with different phases (Grossmann, 2000). This process, paired with a non-linear energy feedback characteristic of turbulence should allow to regenerate the transiently-growing disturbances if Re is sufficiently large (Afshordi et al., 2005). As the strength of the transient amplification was found to have a power-law behaviour as a function of the Reynolds number ($\sim \mathrm{Re}^{2/3}$) and discs possess very large Reynolds numbers ($\mathrm{Re} \sim 10^{13}$), various works suggested that this mechanism could be effective in discs (e.g., Chagelishvili et al., 2003; Yecko, 2004; Mukhopadhyay et al., 2005; Afshordi et al., 2005).

The possibility of a hydrodynamic instability triggering a turbulent state in accretion discs was challenged by Balbus et al. (1996) and Hawley et al. (1999). Their 3D non-linear numerical simulations of rotating shear flows were stable to non-linear disturbances and did not produce any transition to a turbulent state; this is attributed to the stabilising effect of the Coriolis force. Their views are reinforced by non-rotating shear layers quickly becoming turbulent in their numerical simulations. While the stabilising effect of the Coriolis force is well documented, laboratory experiments using Taylor-Couette flows seem to point towards the Coriolis force pushing the onset of turbulence to higher Reynolds numbers, rather than suppressing it altogether (e.g. Richard, 2001), although boundary dynamics may play a role in its development (through the introduction of secondary flows such

as the Ekman circulation). Moreover, Longaretti (2002) believes that the lack of transition to turbulence in the non-linear simulations by Balbus et al. (1996) and Hawley et al. (1999) is due to their low resolution, resulting in a low value of the effective Reynolds number (constrained by the arising numerical viscosity), which is a far cry from the real discs' Re $\sim 10^{13}$. Stabilising rotation of an anti-cyclonic nature has also been found to greatly increase the transition Re in Keplerian flows in both experiments (Tillmark and Alfredsson, 1996) and numerical simulations (Lesur and Longaretti, 2005). However, while the transition Re values for Keplerian rotation were estimated to be in the range Re $= 10^{10} - 10^{16}$, so roughly similar to those observed in accretion discs, the resulting maximum α parameter value from the turbulent state obtained by means of a by-pass transition was found by Lesur and Longaretti (2005) to be $\alpha_{\max} = 3.1 \times 10^{-6}$, which is a few orders of magnitude smaller than the value required to explain observed values for disc accretion rates.

It seems clear then that, although in principle a hydrodynamically-induced turbulent state is possible in a Keplerian system, its anti-cyclonic rotation properties hinder such a turbulent state from making a sizeable contribution to the total outward transport of angular momentum observed.

Rossby wave instability

The Rossby Wave Instability (RWI) is often thought of as the equivalent of the Kelvin-Helmholtz instability in the context of differentially rotating discs. To begin with, Lovelace and Hohlfeld (1978) realised the importance of RWI when looking into thin galactic discs, and it was only about twenty years later that it emerged as a candidate for angular momentum transport in thin Keplerian accretion discs, initially only for non-self-gravitating discs (Lovelace et al., 1999; Li et al., 2000).

A necessary criterion for the RWI to take place is the presence of an extremum in the quantity[1] $\mathcal{L}$ which is defined as

$$\mathcal{L} = \frac{\Sigma\Omega}{\kappa^2}\left(p\Sigma^{-\gamma}\right)^{2/\gamma} = \frac{\Sigma S^{2/\gamma}}{2\left(\nabla \times \boldsymbol{v}\right)_z}, \tag{2.22}$$

where p is the 2D pressure, γ is the adiabatic index, $\boldsymbol{v}$ the velocity of the flow and S the specific entropy of the gas. In the eventuality of the entropy playing

[1]The results illustrated in Chapter 4 of this thesis will however show that the mere presence of an extremum in $\mathcal{L}$ is not a sufficient criterion for the triggering of the RWI.

a negligible part in the system dynamics, the quantity $\mathcal{L}$ is inversely proportional to the potential vorticity $\zeta = (\nabla \times \boldsymbol{v})_z / \Sigma$. The growth of the RWI is due to the co-existence of two Rossby waves – these are vorticity waves which propagate on the potential vorticity gradient – at either side of the extremum, as shown in the lower part of Figure 2.3.

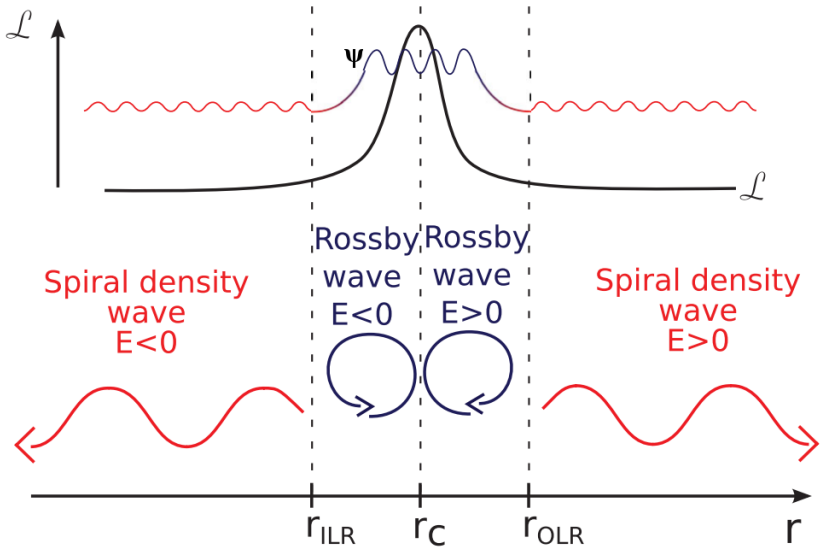


Figure 2.3 – Schematic representation of the Rossby Wave Instability located at a maximum in $\mathcal{L}$ (top half) showing the presence of two Rossby waves propagating either side of the corotation radius r_c; these also emit density waves at either side, which are only allowed to propagate outside the radial locations of the inner and outer Lindblad resonances $r_{\rm ILR}$ and $r_{\rm OLR}$ (Image adapted from Meheut et al. (2013))

The dispersion relation for the Rossby waves is given by

$$\hat{\omega} = -\frac{2\kappa^2 c_s^2}{c_s^2\left(k_r^2 + m^2/r^2\right) + \kappa^2} \frac{m}{r\Sigma} \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}r}, \tag{2.23}$$

where $\hat{\omega} \equiv \omega - m\Omega$ is the wave frequency in the rotating frame, ω is the wave frequency and k_r and m are the radial and azimuthal wave numbers. The wave corotation radius is located where the frequency in the rotating frame $\hat{\omega}$ goes to zero:

$$\hat{\omega}(r_c) \equiv \omega - m\Omega(r_c) = 0; \tag{2.24}$$

this implies, due to the nature of Ω, that $\hat{\omega}(r < r_c) < 0$ and $\hat{\omega}(r > r_c) > 0$. From

the dispersion relation above it is clear that Rossby waves can only propagate if $\mathrm{d}\mathcal{L}/\mathrm{d}r \neq 0$ so they are not allowed at r_c; namely, they are only allowed to propagate in the region where $\mathrm{d}\mathcal{L}/\mathrm{d}r > 0$ if $r < r_c$ and where $\mathrm{d}\mathcal{L}/\mathrm{d}r < 0$ for $r > r_c$.

As shown in Figure 2.3, the Rossby wave located where $r < r_c$ has a pattern frequency which is smaller than the disc's angular frequency ($\omega/m < \Omega$) and therefore is considered to have negative energy (meaning that increasing the amplitude of that wave, energy would be removed from the fluid), while the wave at $r > r_c$ has a pattern frequency larger than Ω ($\omega/m > \Omega$) and carries positive energy. The growth of the instability is caused by the interaction of these two Rossby waves; thanks to their possessing energies of different sign, the total amount of energy can be conserved as the waves grow in amplitude. Their growth mechanism is localised to the region enclosed by the inner and outer Lindblad resonances (r_{ILR} and r_{OLR}), whose location is defined by

$$\hat{\omega}(r_{\mathrm{LR}}) = \pm\kappa(r_{\mathrm{LR}}). \tag{2.25}$$

The differential rotation of the disc couples the Rossby waves to density waves (Tagger, 2001), which are only allowed to propagate outside the locations of the Lindblad resonances; the density waves are instead evanescent for $r_{\mathrm{ILR}} < r < r_{\mathrm{OLR}}$, though quantum tunnelling is possible as shown in the top half of Figure 2.3 by the curve labelled as Ψ.

While the RWI readily develops and grows in the initial conditions (featuring a bump in the pressure) employed by numerical simulations (e.g., Meheut et al., 2010; Meheut et al., 2012a; Meheut et al., 2012b), it is important to consider in what circumstances such conditions may arise in a real disc. Suitable conditions for the RWI to operate have been found to take place at a planet-induced gap's outer edge (e.g., Balmforth and Korycansky, 2001; de Val-Borro et al., 2007), or at the inner edge of a disc's dead zone (e.g., Varnière and Tagger, 2006; Lyra and Mac Low, 2012), with Varnière and Tagger (2006) using a model where the α parameter varies as a function of radius, while Lyra and Mac Low (2012)'s global model includes non-ideal MHD effects and assumes a radially-increasing resistivity. The results from the numerical simulations above seem to point towards the RWI being a potentially interesting instability in the context of producing a turbulent viscosity which matches observation, as long as the conditions required for its onset are met.

The Vertical Shear Instability

Another potential source of hydrodynamic turbulence is the Vertical Shear Instability (VSI) (Nelson et al., 2013); this is closely related to the Goldreich-Schubert-Fricke (GSF) instability and occurs when a vertical shear is present in the system and thermal diffusion negates the stabilising effect of a positive gradient in the entropy, which was discovered during studies of radiative zones in differentially rotating stars (Goldreich and Schubert, 1967; Fricke, 1968). This is a generalised version of the Rayleigh criterion which applies to axisymmetric flows; in the model used, both temperature and midplane density depend on the cylindrical radius r by means of a power law

$$T(r) = T_0 \left(\frac{r}{R_0}\right)^q, \tag{2.26}$$

$$\rho_{\text{mid}}(r) = \rho_0 \left(\frac{r}{R_0}\right)^p, \tag{2.27}$$

where T_0 and ρ_0 are the temperature and midplane density at the fiducial radius R_0 and p and q are real numbers. This relatively simple set-up has the consequence of altering the angular velocity azimuthal profile, which is now no longer constant in z. The angular velocity of the system is now a function of both radius and height from the midplane (z)

$$\Omega(r,z) = \Omega_K(r)\left[(p+q)\left(\frac{H}{r}\right)^2 + (1+q) - \frac{qr}{\sqrt{r^2+z^2}}\right]^{1/2}, \tag{2.28}$$

with $\Omega_K = \sqrt{GM/r^3}$ being the Keplerian profile and H the disc's local scale height.

While a disc with a strictly Keplerian angular velocity profile is hydrodynamically stable according to the Rayleigh criterion

$$\frac{\mathrm{d}j^2}{\mathrm{d}r} > 0, \tag{2.29}$$

where $j = r^2\Omega$ is the specific angular momentum, the stability of a compressible fluid with an angular velocity profile which depends on both radius and height from the midplane – when subject to an axisymmetric, isentropic (i.e., $DS/Dt = 0$) perturbation – is given by the two Solberg-Høiland criteria

$$\frac{1}{r^3}\frac{\partial j^2}{\partial r} + \frac{1}{\rho C_p}(-\nabla P)\cdot\nabla S > 0, \tag{2.30}$$

$$-\frac{\partial P}{\partial z}\left(\frac{\partial j^2}{\partial r}\frac{\partial S}{\partial z}-\frac{\partial j^2}{\partial z}\frac{\partial S}{\partial r}\right)>0, \tag{2.31}$$

with stability obtained if both of the above criteria are fulfilled, where S is the specific entropy and C_p the specific heat at constant pressure.

Accretion discs usually possess $\frac{\partial P}{\partial r}<0$ and $\frac{\partial P}{\partial z}<0$ (for $z>0$), which simplifies the above criteria to

$$\frac{1}{r^3}\frac{\partial j^2}{\partial r}+\frac{1}{\rho C_p}\left(\left|\frac{\partial P}{\partial r}\right|\frac{\partial S}{\partial r}+\left|\frac{\partial P}{\partial z}\right|\frac{\partial S}{\partial z}\right)>0, \tag{2.32}$$

$$\frac{\partial j^2}{\partial r}\frac{\partial S}{\partial z}-\frac{\partial j^2}{\partial z}\frac{\partial S}{\partial r}>0. \tag{2.33}$$

For a disc with a purely Keplerian Ω (and hence j) profile, Equation 2.33 simply reduces to Schwarzschild's convective stability criterion

$$\frac{\partial S}{\partial z}>0; \tag{2.34}$$

if, on the other hand, the angular velocity of the disc considered is also a function of z, a disc which is stable according to Schwarzschild's criterion can nevertheless undergo instability if

$$\frac{\partial j^2}{\partial z}\frac{\partial S}{\partial r}>\frac{\partial j^2}{\partial r}\frac{\partial S}{\partial z}, \tag{2.35}$$

although the condition is not usually met in a standard accretion disc configuration.

If the viscous and thermal evolutions play a part in the evolution of the system then this can no longer be described by the Solberg-Høiland criteria, but instead the stability criteria for the GSF instability must be used. In particular, if the thermal diffusion is much more efficient than viscous diffusion then a fluid element which is perturbed from its equilibrium location would retain its angular momentum, but quickly acquire the entropy of neighbouring elements. This acts against the stabilising effect of a positive entropy gradient described by the Solberg-Høiland criteria, and can cause an instability for a mode that satisfies

$$\frac{\partial j^2}{\partial r}-\frac{k_r}{k_z}\frac{\partial j^2}{\partial z}<0, \tag{2.36}$$

where k_r and k_z are the mode's wavenumbers in the radial and vertical directions, respectively. In a quasi-Keplerian disc, the angular velocity would change more

rapidly across r than z, meaning $|\partial j^2/\partial r| \gg |\partial j^2/\partial z|$, which implies that the modes triggering the GSF instability would have $|k_r| \gg |k_z|$ (i.e., a radial wavelength which is much shorter than that in the z direction).

Numerical simulations find the VSI generates a transport of angular momentum with an α parameter of $\alpha \sim 10^{-3}$ in a PP disc (e.g., Nelson et al., 2013). Because of the efficient thermal diffusion needed and the fact that the introduction of a small, dimensionless kinematic viscosity $\nu > 10^{-6}$ (equivalent to $\alpha \gtrsim 4 \times 10^{-4}$) appears to damp the instability, it is suggested that the VSI likely operates in MRI-stable zones, such as the outer reaches of the dead zone and in the outer parts of PP disc, where the MRI may be quenched by the non-ideal ambipolar diffusion (Armitage, 2011).

Sub-critical Baroclinic Instability

Another promising hydrodynamical instability is the entropy-driven Sub-critical Baroclinic Instability (SBI), initially observed in global numerical simulations of discs possessing a negative entropy gradient in the radial direction by Klahr and Bodenheimer (2003). They observed vortices being formed in the flow as a consequence; the instability was later better defined and characterised using both 2D global (Petersen et al., 2007a; Petersen et al., 2007b) and local (Lesur and Papaloizou, 2010) simulations. These showed that an accretion disc possessing a destabilising entropy gradient which is linearly stabilised by shear can nevertheless undergo a non-linear generation of vorticity through baroclinic interactions.

In order to operate in a self-sustaining fashion, the SBI requires both a radial entropy gradient and thermal diffusion/radiative cooling at a fine-tuned rate which optimises vorticity generation by means of baroclinic interactions (Petersen et al., 2007b; Lesur and Papaloizou, 2010). The first of these conditions is linked to the stability against convective motions of a rotating flow, with the Schwarzschild criterion stating that a flow with no shear is convectively unstable to linear growth of perturbations if

$$N_r^2 < 0, \tag{2.37}$$

with N_r being the Brunt-Väisälä frequency and being defined as

$$N_r^2 = -\frac{1}{\gamma\rho}\frac{\mathrm{d}P}{\mathrm{d}r}\frac{\mathrm{d}}{\mathrm{d}r}\ln\left(\frac{P}{\rho^\gamma}\right) = -\frac{1}{\rho C_p}\nabla P \cdot \nabla S. \tag{2.38}$$

The convective instability criterion for axisymmetric perturbations is altered with the introduction of shear to the Solberg-Høiland criterion

$$N_r^2 + \Omega^2 < 0, \tag{2.39}$$

which is equivalent to Equation 2.30 in the case of a Keplerian Ω profile. The SBI is triggered when the flow is radially unstable according to the Schwarzschild criterion, but Solberg-Høiland stable, in order to avoid the onset of a linear instability, instead.

However the aforementioned simulations show that, although the SBI-driven vorticity develops regardless of the thermal physics applied to the simulated disc, the formed vortices decay on a finite timescale for an unsuitable rate of thermal diffusion or radiative cooling. The SBI can therefore only operate in a self-sustaining way if the thermal diffusion/radiative cooling operates on a fine-tuned rate (with a timescale of $\sim 10\Omega^{-1}$ for Keplerian discs according to Lesur and Papaloizou (2010)), so that the action of buoyancy balances that of dissipative phenomena, as illustrated in Figure 2.4. This ensures that the baroclinic term $\nabla P \times \nabla \rho$ in the vorticity equation, which constitutes both a source and a sink of vorticity, is maintained.

In 2D simulations, the outward transport of angular momentum is – just like in the case of the RWI – caused by shocks of shearing density waves which are generated by the vortices (Johnson and Gammie, 2005). Current estimates for the resulting value of the α parameter due to the SBI are in the order of $\alpha \sim 10^{-3}$, although it is still unclear what the outcome of the SBI is in a 3D simulation, as vertical stratification – as well as radial – is bound to bring new effects (Barranco and Marcus, 2005).

Thermal convection

The idea that vertical convection might be the cause of the anomalous viscosity (and therefore angular momentum transport) in discs has been discussed for a long time (e.g., Cameron, 1978; Lin and Papaloizou, 1980; Lin and Papaloizou, 1985; Ruden and Lin, 1986; Ruden and Pollack, 1991), with a particular emphasis put on the convective phenomenon before the MRI was found to be the source of MHD turbulence.

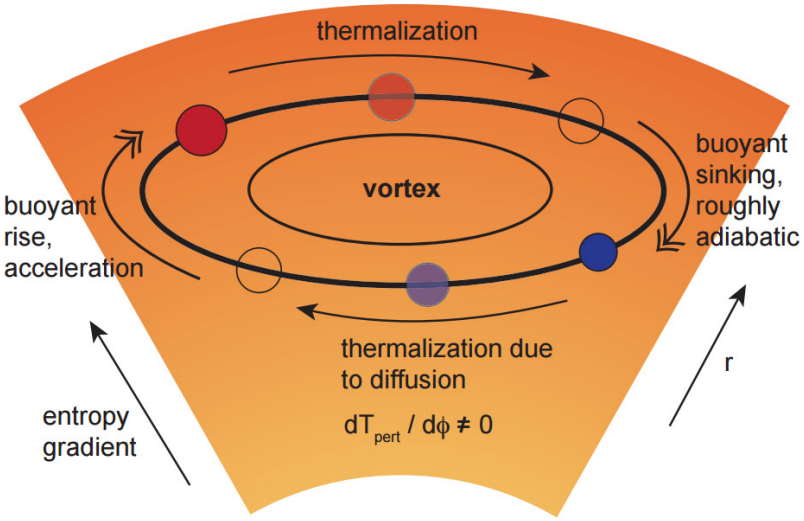


Figure 2.4 – Illustration of the self-sustenance of a SBI vortex in a disk with a radial entropy profile. As fluid elements move radially inward (outward) within the disc due to the action of buoyancy, they will be cooler (hotter) than other surrounding elements. During their azimuthal drift within the vortex, thermal diffusion or radiative cooling act to balance their temperature with that of the background flow, until a thermal equilibrium is reached. This means that within the vortex there exists an azimuthal temperature perturbation ($\mathrm{d}T_{\mathrm{pert}}/\mathrm{d}\phi \neq 0$), which ensures the baroclinic term $\nabla P \times \nabla \rho$ in the vorticity equation, which acts as a source and sink of vorticity, remains non-zero. (Image from Armitage (2011))

In these early works, it was pinpointed that Schwarzschild's criterion for (vertical) convective stability could be violated if the opacity of the disc gas had a particular dependence on temperature. The more recent work by Rafikov (2007), making use of the opacity law (Bell and Lin, 1994)

$$\kappa_R = \kappa_0 P^a T^b$$

and an analysis similar to Lin and Papaloizou (1980), showed that the Schwarzschild criterion for convective instability in optically thick discs (under certain assumptions which are usually met in PP discs) can be equivalently expressed in terms of the power law exponents a and b through

$$\nabla_0 \equiv \frac{1+a}{4-b} \geq \nabla_{\mathrm{ad}}, \tag{2.40}$$

where

$$\nabla_{\rm ad} \equiv \left(\frac{{\rm d}\ln T}{{\rm d}\ln P}\right)_{\rm ad} = \frac{\gamma - 1}{\gamma} \tag{2.41}$$

represents the adiabatic temperature gradient, its value for molecular hydrogen ($\gamma = 7/5$) being $\nabla_{\rm ad} = 2/7$.

Observations in the sub-millimetre part of the spectrum indicate that the opacity due to dust grains at $T \lesssim 100\,{\rm K}$ is well represented by power law exponent values $a = 0$, $b \approx 1$ (Kitamura et al., 2002); this leads to $\nabla_0 \approx 1/3$ and to a convective regime. For $T \lesssim 150\,{\rm K}$ the opacity is dominated by ice grains and well described by $a = 0$ and $b \approx 2$, which yield $\nabla_0 \approx 0.5$, which is again convectively unstable. For T high enough so that ice grains sublimate while silicate and metal grains do not, the latter dominate the opacity law which has $a = 0$, $b \approx 0.5$, yielding again a convectively unstable disc with $\nabla_0 \approx \nabla_{\rm ad}$. Finally in the very high temperature regime ($T \sim 10^3\,{\rm K}$) where all types of dust grains have sublimated, the opacity receives a pressure dependence (i.e., $a \neq 0$) which was missing in the previous cases; this yields $a \approx 2/3$, $b \approx 7/3$ and $\nabla_0 \approx 1$, resulting again in a convectively unstable disc. It is therefore clear that a PP disc can potentially be convectively unstable for a broad range of disc temperatures.

Convective disc models developed by the early simulations mentioned above, such as Lin and Papaloizou (1980), considered convective cells of characteristic velocity $v_{\rm conv}$ and length scale $l_{\rm conv} \sim v_{\rm conv}/\Omega$ and assumed that the kinematic viscosity resulting from their angular momentum transport would be $\nu = v_{\rm conv}^2/\Omega$. The corresponding value of the α parameter associated with thermal convective motions was found to be roughly consistent with observed accretion rates. The linear analysis by Ruden et al. (1988), focused on axisymmetric convective modes, found the transport of angular momentum due to thermal convection to produce $\alpha \sim 10^{-3} - 10^{-2}$. Later works however found that the non-linear evolution of these modes actually resulted in an inward transport of angular momentum, rather than outward (e.g., Kley et al., 1993; Rüdiger et al., 2002).

The interest in convection-driven angular momentum transport in discs has recently revived thanks to the ability to readily perform higher resolution numerical simulations. In particular, the Boussinesq simulation carried out by Lesur and Ogilvie (2010) shows that if the Rayleigh number (which is essentially a ratio of buoyant to dissipative effects) is high enough ($\sim 10^5 - 10^8$), then the direction of angular momentum transport due to convection is reversed to the outward di-

rection. The α parameter is however somewhat small, with its value found to be $\alpha \sim 10^{-4} - 10^{-3}$. The change of sign for the angular momentum transport was attributed to the non-axisymmetric nature of the ensuing turbulent state, which was however missed by earlier works due to their insufficiently high Rayleigh (or Reynolds) number range; this was confirmed by other simulations exploring a similar Rayleigh number range (e.g., Käpylä et al., 2010). Lesur and Ogilvie (2010) however found the angular momentum transport to be ~ 20 times weaker than the turbulent heat transport; this appears to indicate that convective motions are an inefficient means of transporting angular momentum within a disc.

Planet-driven disc evolution

This possible mode of angular momentum transfer, which is a potential alternative only in PP discs as it requires the formation of planets, is based on the fact that the theoretical timescale required to form the core of a giant planet through the core accretion theory and the observed lifetime of PP discs have very similar values. While most works in the literature treat this as a mere coincidence, some have hypothesised that this link is due to planet formation itself playing an active, primary role in the transport of angular momentum within the disc (Larson, 1989; Goodman and Rafikov, 2001; Sari and Goldreich, 2004). Gravitational disc-planet torques exerted at the Lindblad resonance radii would cause angular momentum transfer from orbits interior to the planet's to exterior ones, thereby altering the density profile of the disc and allowing mass to drift towards the central star.

However, as found by Lubow and Ida (2010) when analysing the dynamics of Type I migration, the net torque exerted on the planet (considering both Lindblad and co-rotation torques) is an appreciable fraction (proportional to $\sim H/R$) of the inward torque. This implies that a planet would only induce the accretion of an amount of disc gas mass of the order of its own mass before either migrating into the central star or to the outer part of the disc; this means that low-mass planets cannot be a primary catalyst for disc accretion as hypothesised by Goodman and Rafikov (2001) and Sari and Goldreich (2004), although giant planets might still play a somewhat secondary role in the evolution of the PP disc.

Gravitational instability

An additional effect that needs to be considered in the dynamics of an accretion disc is that due to the disc's own self-gravity. Often discs are assumed to possess little mass compared to the central object and self-gravity effects are neglected; however, in certain conditions (e.g. AGN discs or at early times during the evolution of PP discs when a negligible amount of mass has been accreted or photoevaporated) the disc's self-gravity needs to be taken into account. Disc flows are found to undergo a self-gravitational instability if their Toomre parameter Q, which represents the inverse strength of self-gravity and is given by

$$Q \equiv \frac{c_s \kappa}{\pi G \Sigma}, \tag{2.42}$$

is sufficiently small, where c_s is the sound speed, G the gravitational constant, Σ the disc surface density and κ the epicyclic frequency.

The onset of gravitational instability in discs is due to density enhancements (spiral density waves (SDWs)) and its physical principles are very similar to those of the Jeans instability in the context of a homogeneous fluid (eg Binney and Tremaine, 1987): gravitational forces act on any density enhancement in the fluid by attracting matter to it in a runaway fashion, while pressure forces resist this clumping by trying to pull the matter enhancement apart. There exists a critical size of the initial density enhancement for which the two forces would cancel each other out, called the Jeans Length. If the size of the overdensity is larger than this, gravitational forces are too strong to be suppressed by pressure effects and gravitational instability (or collapse) takes place; similarly, if the initial overdensity is smaller than the Jeans Length, pressure forces are strong enough to overcome gravitational interactions and the initial perturbation propagates through the fluid as a sound wave, although gravity reduces its phase speed. In the case of an astrophysical disc, the additional effect of the rotation must be taken into account, as well as flat geometry if a thin disc is considered.

A simple way to analyse gravitational instability in discs is to assume that the perturbations in question are axisymmetric in nature (i.e. have azimuthal wavenumber $k_y = 0$) and have a radial wavelength $\lambda_{\rm pert}$ which is much smaller than the disc radius, so they can be analysed locally rather than globally. In such a tight-winding approximation ($r/\lambda_{\rm pert} \gg 1$) of a razor-thin (hence 2D) disc, the local dispersion for density waves is given by (Lin and Shu, 1964)

$$\omega^2 = c_s^2 k^2 - 2\pi G\Sigma |k| + \kappa^2, \tag{2.43}$$

where ω is the wave frequency and $k = k_x$ the radial wavenumber. The above expression can be adapted to a disc of finite thickness by introducing a softening length, as discussed in Section 3.3.5. The above dispersion relation includes stabilising effects (such as pressure effects and disc rotation in the first and third terms), as well as the de-stabilising effect of self-gravity (second term). Therefore density waves are unstable when $\omega^2 < 0$, as gravitational interactions dominate, and an exponential gravitational instability is triggered; if $\omega^2 > 0$, on the other hand, density waves are oscillatory in nature and are free to propagate through the disc. It can also be inferred that the stabilising effects of thermal pressure and disc rotation are effective at short and long wavelengths, respectively (large and small k), while self-gravity is most efficient in the intermediate wavelength regime. The simple quadratic form in k of the dispersion relation permits us to identify the wavelength/wavenumber values for which the gravitational instability would take place. It can be easily shown that if $Q < 1$ then the disc would be subject to axisymmetric gravitational instability for some wavenumber range. Representing the inverse strength of self-gravitational effects, the dimensionless Toomre parameter Q (Toomre, 1964) is a key parameter in the gravitational stability of astrophysical discs. Its proportionality to the sound speed c_s, which is itself proportional to $\sqrt{T}$, implies that cooling and heating effects play a crucial role in triggering or stabilising the gravitational instability, as seen in more detail in Section 2.2.2.

The largest growth rate in the unstable k range for $Q < 1$ would coincide with the minimum in ω^2, which can be easily obtained by differentiating Equation 2.43 with respect to k and setting the result equal to zero. This approach indicates that the gravitational instability attains its largest growth rate at

$$k_{\min} = \frac{\pi G\Sigma}{c_s^2}. \tag{2.44}$$

If $Q > 1$, on the other hand, ω^2 is positive for all values of k and the system is therefore gravitationally stable against any axisymmetric perturbation.

The picture is somewhat different for non-axisymmetric disturbances ($k_y \neq 0$, SDWs), which global simulations have shown to be unstable for $Q \lesssim Q_{\rm crit} \sim 1.5$ with multi-armed SDWs emerging in the surface density of the disc and growing on an orbital timescale in the linear case (e.g. Durisen et al., 2007, and references

therein). SDWs are believed to grow through the 'swing' amplification taking place when a wave changes from leading to trailing due to the shear in the system (e.g. Goldreich and Lynden-Bell, 1965a; Goldreich and Tremaine, 1978; Toomre, 1981; Sellwood and Carlberg, 1984). The resulting amplified trailing waves therefore lead to an efficient transport of angular momentum in the outward direction via gravitational and hydrodynamical stresses (Lynden-Bell and Kalnajs, 1972; Papaloizou and Savonije, 1991; Laughlin and Bodenheimer, 1994). If $Q > 1$ the shearing sheet model employed is stable to axisymmetric disturbances and only permits transiently growing non-axisymmetric shearing waves; in the linear regime these transient growths are however quenched by viscosity, meaning the model is linearly stable to non-axisymmetric disturbances. If these transient growths are however coupled to a non-linear positive feedback they can potentially be continuously regenerated, drawing energy from the background flow and leading to the emergence of a self-sustaining turbulent state.

Gravitational instability triggers and the importance of cooling

There exist different mechanisms that can trigger the onset of a gravitational instability; among these are the infall of a clump of material on a disc (Boss, 1997; Boss, 1998), slow infall of material, accumulation of material in the dead zone, close encounters with binary companions or other celestial bodies (Boffin et al., 1998; Lin et al., 1998) and the disc's efficient cooling rate. Particular emphasis is given to the last of these mechanisms, as it is the only one that can be implemented in the razor-thin shearing sheet model used throughout this thesis.

Cooling processes within the disc are mostly responsible for its non-linear development. This is because the Toomre parameter Q can only vary by means of mass redistribution or heating/cooling effects; changes in the surface density would take place on a viscous timescale which, according to the α-disc model, can be expressed as $\sim (r/H)^2(\alpha\Omega)^{-1}$, while changes in the disc temperature occur on a timescale $\sim (\alpha\Omega)^{-1}$ (Gammie, 2001). In the context of a geometrically thin disc, where the aspect ratio of the disc $H/r \ll 1$, gravitational instability is therefore driven and controlled by cooling or heating.

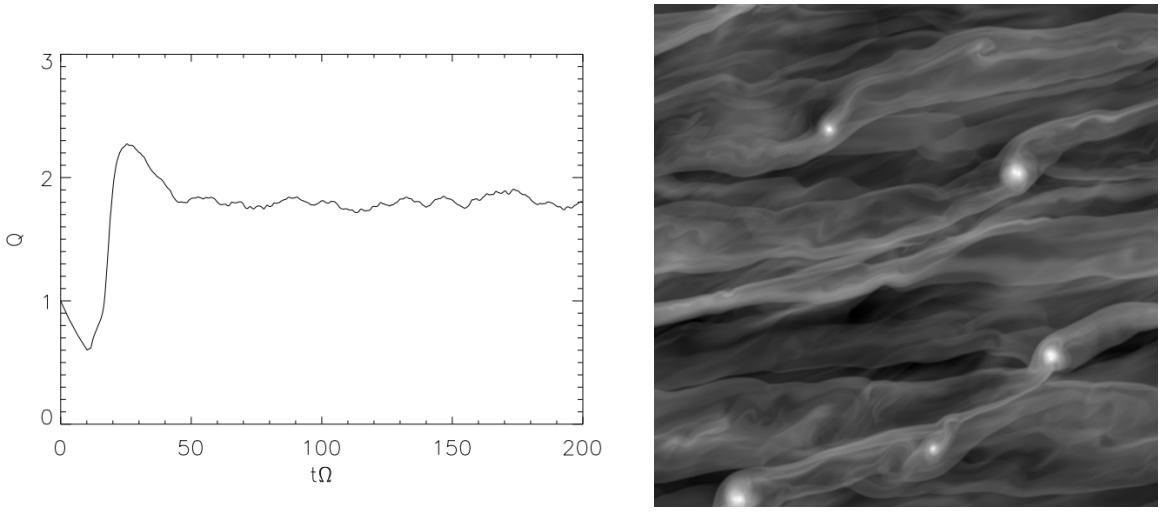
If therefore a disc finds itself in a state of negligible self-gravity (SG), with $Q \gg 1$, the heat produced by SDWs steepening into shocks or by viscous forces trying to smoothen the shocks (we neglect any possible external irradiation) would be negligible; in this case, the disc is able to radiate heat away and lower its temper-

ature, reaching a lower Q. For slow/moderate cooling, as the value of the Toomre parameter is gradually decreased, gravitational effects become more important and the amount of heating produced by the mechanisms mentioned above increases; this means that at some critical value of Q the heating processes would dominate over the cooling, and the disc would warm up again resulting in the gravitational instability (GI) being quenched. It is therefore possible to appreciate that the system is potentially able to enter a self-regulated thermally balanced state by acting like a thermostat; in this scenario the disc would settle into a quasi-steady turbulent state with $Q \approx Q_{\rm crit} \sim 1$ as shown in Figure 2.5a, with growing modes saturating rather than growing exponentially. If the cooling is sufficiently fast, this thermal balance is not achievable as the pressure forces cannot balance the action of SG quickly enough; this results in the denser parts of SDWs eventually fragmenting into bound clumps when the gravitational binding energy exceeds the clump's rotational and internal energies, as seen in Figure 2.5b. As changes in the disc temperature occur on a timescale that's of order the disc's dynamical timescale ($\tau_{\rm dyn} = \Omega^{-1}$, which is the characteristic timescale of GI), the self-regulated state can only arise if the cooling is less efficient than this timescale

$$\tau_c \Omega = \beta \gtrsim 1, \tag{2.45}$$

where τ_c is the cooling timescale and β is a dimensionless constant. The first work to explore the importance of the role of the parameter β in the possible existence of a self-regulating state was Gammie (2001), who first introduced an explicit cooling term in a SG shearing sheet calculation. Gammie (2001) finds the critical β value to be $\beta \sim 3$, with less efficient cooling enabling the disc to settle in a gravito-turbulent state, while more efficient cooling pushes it into a fragmentation regime. The true critical value of β is however still debated, as more recent simulations (e.g. Rice et al., 2003; Rice et al., 2005) have encountered fragmentation events for slightly higher values of β. 2D simulations carried out by Paardekooper (2012) instead found that fragmentation behaves stochastically, with seemingly steady turbulent states being potentially subject to fragmentation up to $\beta \approx 20$, with fragmentation simply becoming less likely the higher the value of β. While the β-cooling prescription has now become commonplace in self-gravitating numerical simulations, it is important to remember that it is only valid for optically thin gas, as it does not take into account optical depth effects (Kratter and Lodato, 2016).

It is also important to highlight that, while this simplistic cooling law represents



(a) **(b)**

Figure 2.5 – Different outcomes from a system becoming gravitationally unstable. In **(a)** the cooling is moderate, which allows the disc to settle into a gravito-turbulent state with $Q \sim 2$; in **(b)**, the cooling applied to the disc was too efficient to allow the self-regulating state to exist, and fragments have formed in the disc's surface density (Figures from Rice et al. (2011)).

an efficient and easy tool to understand the basics of the disc thermodynamic development, a more comprehensive approach should be employed for a more detailed analysis. This would include a more realistic (i.e. non-constant) radiative cooling, as well as potentially adding convective cooling, an external irradiation source and an analysis of how the heat is transported within the disc to its surface. Examples of more thorough thermodynamics can be found in numerical simulations and analytical works; an example of these is Durisen et al. (2007), which found that a drop in the radiative cooling time can correspond to an increase in the disc temperature, which would aid disc stability. This means that a short cooling time is no longer a sufficient condition for the triggering of fragmentation. The detailed studies by Rafikov (2005) and Rafikov (2007) into the dynamics of radiative cooling PP discs – which also include the vertical heat transport due to convection – restrict the fragmentation operation range to $r \gtrsim 100$ AU (for $\sim 0.1 M_{\odot}$ disc around a Sun-like star) as at inner radii longer cooling times make it unlikely (though possible, also based on parameter values); this is in broad agreement with several numerical simulations (eg Nelson et al., 2000; Boley et al., 2006; Cai et al., 2006; Whitworth and Stamatellos, 2006; Boley, 2009; Forgan et al., 2009; Cossins et al.,

2010). Rafikov (2007) also shows that the action of convection for inner radii is not sufficient to lower cooling times enough to trigger fragmentation, contrary to the claims of Boss (2004), Mayer et al. (2007) and others. This is because, although convection can transfer heat from the midplane to the radiative atmosphere on a dynamical timescale, the subsequent heat dispersal into space is limited by the adiabatic temperature gradient. Lastly, fragmentation at radii of at least a few tens of AU are boosted in the early phases of PP discs, when they are at their most massive and extended configuration and accretion from their parent cloud is still ongoing (Vorobyov and Basu, 2006; Clarke, 2009; Boley, 2009; Boley et al., 2010; Vorobyov and Basu, 2010).

Angular momentum transport by GI

The nature of the angular momentum transport triggered by non-axisymmetric gravitational instability has been a topic of debate in the literature, as the local model of the shearing sheet was attempting to replicate a phenomenon (self-gravity) which is usually described globally. Several numerical simulations (e.g. Laughlin and Bodenheimer, 1994; Laughlin and Rozyczka, 1996; Gammie, 2001; Lodato and Rice, 2004; Lodato and Rice, 2005; Mejía et al., 2005; Cossins et al., 2009) have looked at the issue, as well as to whether the α prescription can effectively represent GI. The results from such numerical simulations appear to show that in massive discs with $M_{\rm disc}/M_c \gtrsim 0.25$ (M_c being the mass of the central object) GI behaves globally, due to the domination of low k_y SDW modes; this means that GI in massive discs cannot be represented by the α model. Moreover in this particular scenario GI occurs in strong transient episodes, making the gravito-turbulent state difficult to obtain (Lodato and Rice, 2005; Lodato, 2007), which produce very large torques and angular momentum transport ($\alpha \sim 0.5$).

On the other hand, if the disc matches the condition $M_{\rm disc}/M_c \lesssim 0.25$, its quasi-steady gravito-turbulent state presents several higher-k_y SDW modes; these are thought to interact with each other, dissipate and shock, causing their propagation to be contained within a small radial range. This means that angular momentum transport induced by GI in low-mass discs can be described locally (Gammie, 2001; Lodato, 2007) and can be well represented using the α model formalism. In this scenario the value of α, which is a combination of gravitational and hydrodynamical stresses as shown in Section 6.3.1, is given by

$$\alpha = \frac{1}{\gamma(\gamma-1)\Omega\tau_c}\left|\frac{\mathrm{d}\ln\Omega}{\mathrm{d}\ln r}\right|^{-2} = \frac{1}{q^2\gamma(\gamma-1)\Omega\tau_c}. \tag{2.46}$$

As it can be seen, the only factor that affects the angular momentum transport rate α (for a fixed adiabatic index γ) is the cooling time τ_c given by Equation 2.45. This means that if the system is able to settle into a self-sustained gravito-turbulent state, the resulting angular momentum transport should be independent of time and therefore constant. This is the local analogue of the result stating that a disc's steady-state surface brightness is independent on the details of the mechanism triggering the angular momentum transport (Pringle, 1981; Balbus, 2003).

Evidence of GI in discs

While the following chapters of the thesis will focus on an academic discussion of the complex dynamics of self-gravitating discs, a preamble regarding how SG ties in with observations of real discs is important.

Self-gravitational interactions become of importance when the disc mass is a significant fraction of the central object's mass; a simple estimate can be made by rewriting the Toomre parameter – which quantifies the strength of the self-gravity interactions in the disc – in terms of observable parameters as given below (Kratter and Lodato, 2016)

$$Q = \frac{c_s\Omega}{\pi G\Sigma} = f\frac{M_c}{M_{\mathrm{disc}}}\frac{H}{r}, \tag{2.47}$$

where M_c and M_{disc} are the masses of the central object and of the surrounding disc, H is the disc's scale height and r its radius and f is a numerical factor of order unity. Although the dependence of the disc aspect ratio factor H/r on the disc temperature (through the sound speed) prohibits a fixed M_{disc} threshold, it is however possible to carry out an estimate for the disc mass required to trigger self-gravity.

For protoplanetary discs, models typically choose a disc aspect ratio of $H/r \approx 0.05 - 0.1$ (e.g. Lubow and D'Angelo, 2006; Lorén-Aguilar and Bate, 2015); this implies that self-gravitational interactions become of importance for mass ratios in excess of $M_{\mathrm{disc}}/M_c \approx 0.01$. As it can be easily deduced from the equation above, cold thin discs are more likely to be self-gravitationally active. Moreover, as mentioned in Section 1.1, PP discs go through four evolutionary stages; as mass

is gradually accreted onto the protostar and swept away by the stellar wind, the earlier phases of evolution would be the ones more susceptible to the onset of gravitational instabilities. Masses inferred from submillimetre fluxes of optically thin dust for Class II systems show that only between $10-20\%$ of targets possess a disc massive enough (compared to the star) to be self-gravitating (Andrews et al., 2013); masses inferred by submillimetre fluxes are however subject to several assumptions and therefore can only provide a lower limit on the value. Younger systems of Classes 0 and I are more difficult to observe due to them being embedded and obscured by the envelope; this causes surveys to encompass smaller samples, potentially resulting in a biased fraction of discs massive enough to be subject to GI. Nevertheless, they find $10-20\%$ of Class I (Esner, 2012; Mann et al., 2015) and roughly 50% of Class 0 (Tobin et al., 2013; Tobin et al., 2015) systems to have mass ratios $M_{\rm disc}/M_c$ large enough for gravitational instability.

AGN discs on the other hand possess a much smaller aspect ratio, with models indicating the value being $H/r \approx 10^{-3}-10^{-2}$ (Collins-Souffrin and Dumont, 1990; Sirko and Goodman, 2003). This has strong implications for self-gravity, as shown by Equation 2.47, as a much smaller mass ratio $M_{\rm disc}/M_c$ is necessary for SG to be important in the dynamics of the system. Indeed, gravitational instability can take place in AGN discs with mass ratios $M_{\rm disc}/M_c \approx 10^{-3}$. This value can be used to make an estimate of the radius at which the disc would start to be gravitationally unstable (Lodato and Bertin, 2003); this is found to be $r \sim 10^3 R_S$, where R_S is the Schwarzschild radius, or $r \approx 10^{-2}$pc for a black hole of mass $10^8 M_\odot$. Spectral emissions from AGNs, on the other hand, suggest discs around black holes can extend up to radii of the order of 1pc (Greenhill and Gwinn, 1997; Kondratko et al., 2006, e.g.), or $r \sim 10^5 R_S$ for a $10^8 M_\odot$ black hole, although models are not in agreement regarding the disc size (Thompson et al., 2005; King and Pringle, 2007).

GI in galactic discs

Galaxies are categorised according to their properties, such as the presence/absence of a central bar or the presence and number of spiral arms. These are thought to be quasi-stationary density waves (Lin and Shu, 1964) obeying the dispersion relation given in Equation 2.43, and are amplified by the stellar component of self-gravity.

Another process in which self-gravity plays a key role in galactic discs is the so-called feathering (Lynds, 1970; Elmegreen, 1980). This represents density fluctuations, seemingly spaced by a roughly regular separation, fanning out of the

galactic spiral arms. The phenomenon is widely observed in optical and infrared wavelengths, and is caused by the self-gravitational contraction of gas over-densities in the spiral arms following shocks (Shu et al., 2007; Lee and Shu, 2013).

GI in protoplanetary discs

Like in galactic discs, self-gravity also produces spiral density waves in protoplanetary discs, which have recently been linked with planet formation mechanisms (Pérez et al., 2016). Planets are also believed to form through the process of disc fragmentation. Although observational evidence on the role of fragmentation in planet formation is still scarce, studies on directly imaged wide-orbit exoplanets hint towards fragmentation as a likely formation mechanism (Marois et al., 2008).

It is also possible that fragmentation may instead lead to the formation of bodies larger than giant planets, resulting in the creation of brown dwarfs (Lodato et al., 2005; Stamatellos and Whitworth, 2009). Numerical simulations (Bonnell et al., 2008; Stamatellos et al., 2012) and observations (Hinkley et al., 2010; Riaz et al., 2012) both show that the observed properties of many brown dwarfs are consistent with a disc fragmentation formation scenario.

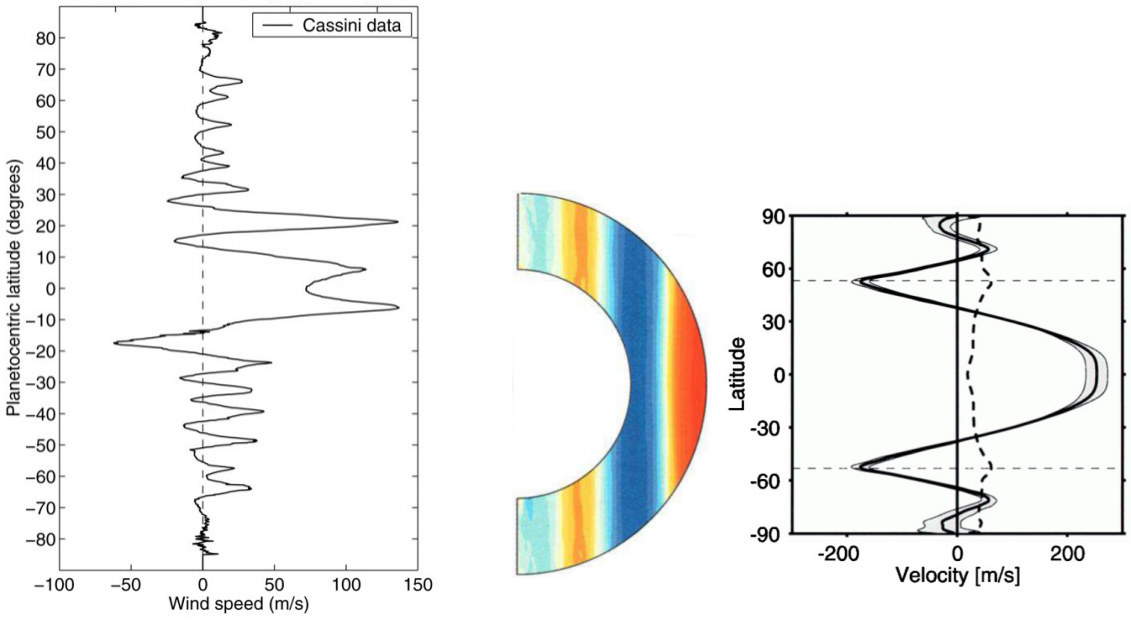
GI in planetary rings

Lastly, the effects of self-gravity can also be observed in planetary rings. These consist of large amounts of particles, typically measuring up to $\sim 10^2 - 10^3$ m, which independently orbit around a planet and are prevented from forming a moon by the tidal torque exerted by said planet. These particles are subject to gravitational forces, which act to clump them together, and to the differential rotation of the ring, which acts to tear apart bound objects. The action of these forces results in the formation of gravitational wakes (Michikoshi et al., 2015); these do not represent aerodynamical wakes, as the turbulent motions of a fluid after an object has moved through it, but non-axisymmetric spiral density structures. These wakes, not to be confused with similar non-axisymmetric structures created by the presence of moons in the rings, have been observed in Saturn's rings (e.g. Salo, 1992; Colwell et al., 2006; Hedman et al., 2007) and thoroughly studied with N-body simulations (e.g. Michikoshi et al., 2015).

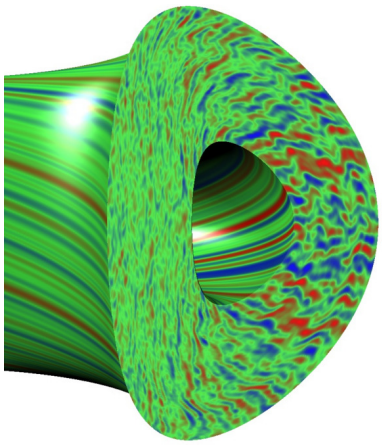
2.3 Zonal flows

Zonal flows are axisymmetric shear flows which are characterised by a recurring alternating pattern, and occur in a multitude of physical situations. Here on Earth zonal flows play a central role in the atmospheric weather dynamics, creating sharp temperature gradients that trigger the formation of cyclones, bringing cloud formation and precipitation to varying degrees. Similar phenomena also occur in the atmosphere of the other planets of the Solar System; gas and ice giants represent one of the best ways to study zonal flows thanks to their weather being particularly extreme. The most clear example of zonal flows in the Solar System is Jupiter's famous latitudinally striped pattern of highly axisymmetric bands – having a width of $4000 - 10000$ km (Gierasch et al., 1979) – which are locally distorted by disturbances in the order of $\sim 100 - 1000$ km and also present jets.

Jupiter's zonal flow profile – believed to potentially be a manifestation of phenomena taking place in the lower strata of its atmosphere – has been measured by tracking cloud motions (Limaye, 1986), as shown in Figure 2.6a. This displays the mean zonal flow velocity as measured by the Cassini spacecraft, whose magnitude is found to dominate over north-south motions and to be at its strongest around the planet's equator. Although accurate measurements have been conducted as early as a few decades ago, the origin of the Jovian characteristic bands is still controversial. It was first thought that the observed zonal flow could be a reflection of the dynamics within a shallow layer in the upper part of the planetary atmosphere driven by solar heating (Williams, 1985). The discovery made by the Galileo probe that Jupiter's mean zonal flow speed increases with depth (Atkinson et al., 1996) appears to suggest that the dynamics causing the zonal flow are however internally driven. The focus therefore shifted towards convection within the molecular hydrogen layer, which takes up a sizeable fraction of both Jupiter's and Saturn's atmospheres (Guillot, 1999); simulations looking at the effects of convection within an atmospheric shell have indeed given encouraging results (Sun et al., 1993; Christensen, 2001; Heimpel and Aurnou, 2007), such as Figure 2.6b; this shows the mean zonal velocity on the cross-section of the employed convective shell as well as the zonal velocity profile as a function of latitude (full line). This features a peak in the velocity magnitude around the planetary equator – which is a feature also observed in the Solar System planets, such as Jupiter as shown in Figure 2.6a – although there appear to be fewer bands than on the Jovian "surface". The number of the



(a) (b)



(c)

Figure 2.6 – Examples of zonal flows: **(a)** Jupiter's average zonal mean flow as measured by Cassini (Figure adapted from Heimpel and Aurnou (2007)); **(b)** zonal flow velocity cross-section and profile encountered in a simulation looking at convection in a rotating spherical shell in a planetary-like atmosphere (Image adapted from Christensen (2001)); **(c)** output from a turbulent dynamics plasma simulation showing the presence of a zonal flow in the torus (Image from the GYRO simulation, Waltz et al. (2002)).

zonal bands can however be altered by varying the thickness of the convective shell (Christensen, 2001).

The presence and development of zonal flows have also been reported in laboratory plasma experiments and simulations (e.g. Waltz et al., 2002; Diamond et al., 2005) looking into the phenomenon of drift wave turbulence, of which zonal flows are presently believed to be an intrinsic player, particularly in the role of turbulent self-regulation. Figure 2.6c shows the output of a tokamak (a magnetic confinement device that constrains the plasma to follow a torus shape) plasma numerical simulation by the GYRO code (Waltz et al., 2002), showcasing a zonal flow pattern in the resulting torus.

Recent numerical works focusing on the magnetohydrodynamic (MHD) turbulence in shearing sheet model accretion discs have also produced zonal flows (e.g. Johansen et al., 2009; Simon et al., 2012; Kunz and Lesur, 2013; Bai and Stone, 2014). These appear to become more prominent and less intermittent as the box size is increased, filling the radial extent of the domain, and becoming more long lived and less prone to radial drift (Bai and Stone, 2014). Their amplitudes also appear to increase when considering larger boxes, and they usually possess the largest resolved radial wavelength (i.e. $k_x = \frac{2\pi}{L_x}$, where k_x and L_x are the radial wavenumber and box size) (Johansen et al., 2009; Simon et al., 2012). Lastly, the zonal flows observed in such simulations do not appear to be caused by specific initial conditions applied to the system, as the work by Simon et al. (2012) was initialised with either a uniform toroidal field or two flux tubes and found no significant difference in the emergence time of the axisymmetric structure, nor in its properties.

The 2D shearing sheet model calculations of an incompressible, inviscid hydrodynamical disc by Lithwick (2007) and Lithwick (2009) found that zonal flows can break up into vortices (see also Gill, 1965; Lerner and Knobloch, 1988) due to the action of the Kelvin-Helmholtz (KH) instability. The resulting vortices are found to be located only in the zonal flow bands whose vorticity matches the sign of the background value, as bands with vorticity of opposing signs are not unstable. The non-linear instability pinpointed by Lithwick (2007) and Lithwick (2009) is however somewhat elusive, as it acts at very long azimuthal wavelengths. This implies that in order for the instability to be detected, the simulation must have a sufficiently large domain in y to resolve small values of the azimuthal wavenumber $|k_y|$ which satisfy

$$|k_y| \lesssim \frac{1}{q}\frac{|\omega|}{\Delta x_{\rm zf}}, \tag{2.48}$$

where $|\omega| \lesssim q$ and $\Delta x_{\rm zf}$ are the vorticity excess and width of a zonal flow band. It can be noticed that the zonal flow band in question is theoretically always unstable to perturbations having a sufficiently long wavelength, no matter the values of $|\omega|$ and $\Delta x_{\rm zf}$, even if the amplitude (i.e. $|\omega|$) of the zonal flow is infinitesimally small. The easiest way to detect this non-axisymmetric instability is therefore to consider a computational domain that is significantly stretched in the azimuthal direction, provided the zonal flow strength is sufficient to produce an instability with a substantial growth rate so that the system undergoes the instability regardless of the stabilising effect from any viscosity (explicit or numerical). It is also crucial that the incompressible conditions outlined in Lithwick (2007) and Lithwick (2009) are not violated on such elongated boxes. Other 2D numerical simulations of discs found that coherent vortices caused by the presence of an axisymmetric structure can have a long decay time (Umurhan and Regev, 2004; Johnson and Gammie, 2005), which scales as $t_{\rm decay} \sim {\rm Re}^{2/3}$ (Yecko, 2004). This is despite the numerical simulations only being able to recreate a disc with a modest Reynolds number ${\rm Re} = VL/\nu$ (where V and L are the disc's characteristic velocity and lengthscale) compared to those observed in real life discs. Indeed, the work by Umurhan and Regev (2004) could only implement Reynolds number values ($\sim 10^5$) which were a few orders of magnitude smaller than typical accretion disc values ($\sim 10^{13}$). The number of coherent vortices picked up by Umurhan and Regev (2004) appears to increase as the box is enlarged, suggesting their origin is intrinsic to the dynamics of the flow, rather than a spurious effect prompted by constraining boundary conditions. These zonal flow-induced vortices also appear to generate a transport of angular momentum in the outward direction, which is required to allow material to accrete onto the central object. It is therefore possible that these vortices could represent a viable alternative to MHD turbulence in those regions of accretion discs where the ionisation is insufficient to trigger the MRI, provided that an appropriate inherent mechanism for the generation of zonal flows is identified.

Zonal flows are also believed to play a key role in the dynamics of planetary formation taking place in protoplanetary discs. In particular, they are seen as a potential way to overcome the 'metre-size barrier'; this is the name given to the severely hampered chances of metre-size bodies growing and surviving within the disc. This is due to various reasons, including a rapid radial inward drift

experienced by the bodies due to their decoupling from the gas (Weidenschilling, 1977), and collisions between roughly metre-sized bodies being often destructive in nature. The presence of a zonal flow can affect the former of these effects by altering the gas drag the bodies experience – which is due to the gas rotating around the protostar somewhat slower than the bodies and the resulting friction removing angular momentum from the metre-sized bodies – either reducing the radial migration (which without the presence of an axisymmetric structure would cause a metre-sized body to spiral into the protostar from a distance of 1 AU on a timescale of $\sim$ 100 years (Weidenschilling, 1977)) or halting it completely (Whipple, 1972; Klahr and Lin, 2001; Haghighipour and Boss, 2003; Fromang and Nelson, 2005; Kato et al., 2009). The slowing down of the inward drift would in turn boost the local concentration of metre-sized bodies, assisting their growth by coagulation and the onset of gravitational instability which, if sufficiently powerful, can lead to the disc fragmenting.

2.4 Aims and outline of the thesis

As mentioned previously, the Shearing Sheet model considered in this thesis is only subject to linear axisymmetric gravitational instability for $Q < 1$; for weaker gravitational conditions, the model only admits non-axisymmetric disturbances in the form of transiently growing shearing waves, which in the linear regime quickly decay due to viscous effects, meaning the model is linearly stable to non-axisymmetric perturbations. However if a coupling exists between the non-axisymmetric shearing waves and a non-linear feedback, energy could be continuously extracted from the background flow to feed these transient growths, ensuring these can be regenerated and that a state of self-sustaining turbulence can exist. 2D shearing sheet numerical simulations (e.g. Gammie, 2001; Johnson and Gammie, 2003; Rice et al., 2011; Shi et al., 2016) of self-gravitating discs found that, for an intermediate range of cooling times, the disc can settle into a state of gravitationally-driven turbulence having an average Q value of $\overline{Q} \approx 2$, which means the disc is stable to axisymmetric perturbations. The situation appears somewhat analogous to the hydrodynamical sub-critical transition to turbulence discussed in Section 2.2.2, with laboratory experiments showing flows becoming turbulent for Reynolds numbers Re smaller than the critical value required to produce a linear instability (Mukhopadhyay et al., 2005).

In this thesis, zonal flows are brought forward and considered as a potential source of feedback on the gravito-turbulent state. Firstly their contribution to the stability of accretion disc flows is analysed with a linear calculation with respect to non-axisymmetric disturbances (Chapter 4); the second linear analysis (Chapter 5) focuses on the stability of zonal flows themselves, considering an extension to the classical theory of thermal and viscous disc instabilities, as well as the possibility of overstability. All the instabilities studied are seen to affect intermediate wavelengths (i.e. $\sim$ a few H), with the non-axisymmetric instability found to play a role in constraining the amplitude and lengthscale of the zonal flows, which is of importance when considering the self-sustenance of the gravito-turbulence state. This is carried out with a non-linear numerical code, entirely developed for the purpose of this thesis, with information from the linear stability calculations also used in order to fully understand the feedback mechanism sustaining the gravito-turbulent regime.

The rest of this thesis is structured as follow: Chapter 3 introduces the non-linear spectral code `CASPER` and its technical specifications, such as the full equations describing the system, viscosities, (horizontal) thermal diffusion, time stepping algorithm and anti-aliasing methods. Several tests of early development stages of the code are also provided where simplified problems are tackled, to assess the numerical soundness of the methods. Due to the habitual difficulty of spectral codes in successfully capturing and resolving shocks, further tests are conducted showing the shock-capturing accuracy of the method employed; such shortcomings of spectral codes are however eclipsed by the beneficial effect of being able to control the amount of viscosity/diffusion present in the system, which play a crucial role in the work carried out in this thesis.

A linear stability calculation considering non-axisymmetric perturbations in ideal conditions (i.e. no cooling or viscous effects) is carried out in Chapter 4, both in isothermal and adiabatic scenarios. The stability of the flow, on which a zonal flow has been imposed, is studied both in the absence and presence of self-gravitational effects. In the case with no self-gravity, the strength of the observed non-axisymmetric instability is studied as a function of both zonal flow amplitude and wavelength, with the instability found to only operate on zonal flows of sufficiently short wavelength. The SG case presents an additional instability, whose properties are studied and compared to those of the non-SG instability previously mentioned. Adiabatic conditions are also considered, with a yet different instabil-

ity found when considering a zonal flow in the entropy but not in the potential vorticity (unlike the instabilities mentioned above).

Chapter 5 presents a linear stability analysis with respect to axisymmetric perturbations, this time conducted entirely in adiabatic conditions, but in this case the disc considered is viscous, cooled on a constant timescale and also presents a horizontal thermal diffusion, while the zonal flow is freely evolving. Although the solutions to the inviscid case with no cooling or thermal diffusion are well known, it is not trivial to pinpoint whether the introduction of these effects would have a stabilising or de-stabilising effect. The stability analysis, which focuses on axisymmetric perturbations of wavelength $\lambda_{\rm pert} \sim H$ and therefore represents an extension to the classical analysis on viscous and thermal instabilities, deals with all four solutions to the system: the two density waves and the non-oscillating structures in potential vorticity and entropy. The instabilities are analysed in parameter space under the assumption of thermal balance, and the growth rates of the PV and entropy modes are found to be significantly affected by the coupling between the two quantities, which is instead negligible in the classical theory.

In Chapter 6 the results obtained using the numerical code `CASPER` are analysed; a zonal flow, which is not initialised in the equations, is seen to emerge and be continuously regenerated as the system settles into a state of gravito-turbulence. One of the axisymmetric instabilities found in Chapter 5 is found to be responsible for the zonal flow regeneration, with the non-axisymmetric gravitational instability studied in Chapter 4 also being involved by constraining the zonal flow properties. Although the latter instability is found to disrupt the zonal flow, the resulting shearing waves are then observed to be the seed for the zonal flow regeneration, also thanks to the shearing nature of the flow. The characteristics of the gravito-turbulence state are subsequently explored as a function of some of the initial parameters.

Chapter 3

The code

As part of this thesis, a non-linear compressible spectral code was developed from scratch using the `C` language. This code, which was dubbed `CASPER`, also includes the effects of self-gravity in order to study the evolution of gravitationally-induced turbulence on an accretion disc flow. The main results obtained using the code are discussed in Chapter 6 of this thesis.

Developing the code to its final stage took a considerable amount of time as it was paramount to ensure that its quality and accuracy were of a high standard; to this end, several tests were carried out in order to check its behaviour with known physical problems before moving on to simulate more complex systems. These are presented in Section 3.5.

The code makes use of the Runge-Kutta 3$^{\text{rd}}$ order method of iteration (Canuto et al., 2006), which was chosen for its near marginal stability at small enough values of the simulation timestep Δt in the case of wave propagation. The linear stability regions are established by considering the simple linear differential equation $\mathrm{d}x/\mathrm{d}t = \lambda x$, where λ is a constant which could be a complex number. Evolving said equation over a single step, it is safe to assume that $x_{n+1} = g x_n$, where $g = g(\lambda\Delta t)$ is the increment function; the form of $g(\lambda\Delta t)$ depends on the iterative method employed. The above iteration is however divergent for $|g(\lambda\Delta t)| > 1$, meaning the iterative method used is stable for $|g(\lambda\Delta t)| \leq 1$. Contour lines for $|g(\lambda\Delta t)| = 1$ can therefore be plotted as a function of $\lambda\Delta t$, as shown in Figure 3.1, to visualise the stability characteristics of various iterative methods. This shows that the RK3 method is marginally stable for imaginary λ for $|\lambda\Delta t| \lesssim 1.75$.

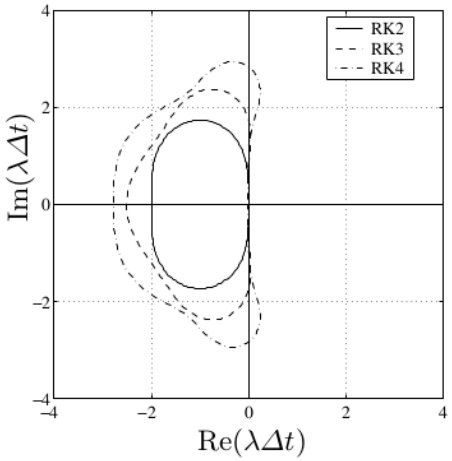


Figure 3.1 – Absolute stability regions for the Runge-Kutta 2nd, 3rd and 4th order methods. Stability is achieved within the regions enclosed by the curves (Image: Canuto et al. (2006)).

3.1 A spectral code

Spectral codes are a class of method used to solve systems of ordinary differential equations (ODEs), partial differential equations (PDEs) and eigenvalue problems in various branches of mathematics. Spectral methods consist in expressing the solutions to such differential equations in the form of a sum of basis (or trial) functions as below:

$$f(\boldsymbol{x}, t) = \sum_{k=0}^{N} \tilde{f}_k(t)\phi_k(\boldsymbol{x}). \tag{3.1}$$

While other methods – such as finite difference or finite element methods – are based upon the same basic assumption, the uniqueness of spectral methods is that it makes use of global trial functions (e.g. Fourier series), rather than local ones.

There exist different types of spectral methods, which are distinguished by the way the expansion coefficients $\tilde{f}_k$ are selected and by how the residual function (an approximate measure of the solution's quality) is minimised; the main types are: Galerkin, Tau-Lanczos, Least squares and collocation (or pseudo-spectral). `CASPER` makes use of the pseudo-spectral approach in which the basis functions are chosen in order to satisfy the imposed boundary conditions and to reduce the residual to zero at all grid points. Pseudo-spectral methods offer a computational advantage over other spectral methods for non-linear conditions with a fine grid;

this is a consequence of their ability to deal with non-linear terms in real space, therefore avoiding a corresponding convolution in Fourier space. This means that for these non-linear terms the pseudo-spectral methods need to compute $\sim N \log N$ operations, rather than $\sim N^2$ (Mainardi and Pagnini, 2002).

A few factors were at the root of the choice of a pseudo-spectral method to tackle this problem. One of these is spectral methods' extremely good general performance when periodic boundary conditions are applied to the system (as true in both dimensions in this case), as the Fourier series and the optimised Fast Fourier Transforms (FFTs) make for an efficient computation. On the other hand, spectral methods are known to struggle with non-periodic boundary conditions, as these can introduce discontinuities, with which the methods struggle to cope. Considering self-gravity in a spectral method is also straightforward, as the linear character of Poisson's equation allows the disc potential to be most easily expressed in Fourier space, as seen later in Equations 3.7–3.8. Furthermore, spectral methods' ability to consider each resolved wavelength separately effectively means they are able evolve each of these modes independently. This represents a very useful feature when one aims to confirm and bridge the results of linear stability analyses (carried out in Chapters 4 and 5) in a full non-linear context (Chan et al., 2006).

Also, the efficiency and precision of spectral methods have made them a particularly useful tool in solving PDEs as they provide a better accuracy than other methods when analysing smooth flows. In fact, the rate of convergence of a smooth solution's errors associated with a spectral method with N gridpoints follows an exponential function ($\propto (1/N)^N$), while as a means of comparison the precision of Finite Element and Finite Difference methods is given by a negative power of N ($\propto (1/N)^p$, where p is the order of the method) (Costa, 2004). This means that a spectral method can achieve a higher accuracy with a much smaller number of grid points, hence resulting in a substantially shorter computational time. An example of this is given in Table 3.1, which lists the errors obtained in the approximation of $\partial \mathrm{e}^{\sin x}/\partial x$ using a Finite Difference method (2^{nd} and 4^{th} order) and a Spectral method. The spectral method presents a greatly reduced error for all values of N analysed; the lack of improvement in the spectral method error when N is increased from 64 to 128 is thought to be due to machine precision having already been reached.

Resolving shocks and discontinuities represents the greatest problem for codes employing a spectral method to solve the equations characterising a transonic flows.

Table 3.1 – Comparison in the errors in the approximation of the first derivative of $e^{\sin x}$ obtained using a Finite Difference method (2^{nd} and 4^{th} order) and a Fourier Spectral collocation method (Costa, 2004). No improvement in the accuracy of the spectral method is observed increasing N from 64 to 128 as the code has reached the maximum machine precision

N	FD 2^{nd} order	FD 4^{th} order	Spectral
16	3.40×10^{-1}	1.38×10^{-1}	4.32×10^{-3}
32	9.36×10^{-2}	1.58×10^{-2}	1.76×10^{-7}
64	2.58×10^{-2}	1.13×10^{-3}	2.39×10^{-14}
128	6.51×10^{-3}	7.59×10^{-5}	7.21×10^{-14}

However spurious effects such as the Gibbs phenomenon can be safely dealt with by applying an explicit viscosity – as done in `CASPER` – as well as employing a hyper viscosity, or an appropriate filtering (Majda et al., 1978; Jou and Mueller, 1985; Gottlieb and Hesthaven, 2000).

Spectral methods are used extensively in fluid dynamics, both in astrophysics and in aerospace engineering. In the astrophysics field, `SNOOPY` (Lesur and Longaretti, 2005) represents a widespread MHD incompressible pseudo-spectral code which is often used to analyse MRI and the turbulent state it induces. Spectral methods have also been used to analyse vortex dynamics in stratified flows (Barranco and Marcus, 2006), MHD stability (Taylor, 1994), black hole evolution (Tichy, 2006) and mergers (Szilágyi et al., 2009), stellar core convection (Cai et al., 2011) and aerodynamics (e.g. Streett, 1987; Ekaterinaris, 2005; Kondaxakis and Tsangaris, 2008) among others.

3.2 The shearing sheet model

The analysis carried out in the `CASPER` code is based on the shearing sheet model first employed by Goldreich and Lynden-Bell (1965b) in the study of unstratified galactic discs. This model is local in nature and uses a Cartesian frame of reference with periodic boundary conditions centred around a fiducial disc radius R_0. This means that the sheet – which has radial and azimuthal dimensions L_x and L_y such that L_x, $L_y \ll R_0$ – corotates with the disc.

In the frame of reference of the sheet, a compressible, viscous fluid is described by the following continuity and momentum equations:

$$\frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma \boldsymbol{v}) = 0, \tag{3.2}$$

$$\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v} \cdot \nabla \boldsymbol{v} + 2\boldsymbol{\Omega} \times \boldsymbol{v} = -\nabla \Phi - \nabla \Phi_{\mathrm{d,m}} - \frac{1}{\Sigma} \nabla P - \frac{1}{\Sigma} \nabla \cdot \boldsymbol{T}, \tag{3.3}$$

where Σ represents the surface density of the disc, $\boldsymbol{v}$ the velocity of the flow, $\boldsymbol{\Omega} = \Omega \boldsymbol{e}_z$ the disc angular velocity, $\boldsymbol{e}_z$ the unit vector parallel to the z-axis, $\Phi_{\mathrm{d,m}}$ the disc potential calculated at its mid-plane, $\Phi = -\Omega^2 q x^2$ the effective potential, $q = -\mathrm{d} \ln \Omega / \mathrm{d} \ln r$ the dimensionless shear rate ($q = 3/2$ for a Keplerian disc), P the 2D pressure and

$$\boldsymbol{T} = 2\mu_s \boldsymbol{S} + \mu_b (\nabla \cdot \boldsymbol{v}) \boldsymbol{I} \tag{3.4}$$

the stress tensor for the bulk and shear dynamic viscosities ($\mu_i = \Sigma \nu_i$), with ν_i being the kinematic viscosity and $\boldsymbol{I}$ the unit tensor, and $\boldsymbol{S}$ being the traceless shear tensor given by

$$\boldsymbol{S} = \frac{1}{2} \left[\nabla \boldsymbol{v} + (\nabla \boldsymbol{v})^T \right] - \frac{1}{3} (\nabla \cdot \boldsymbol{v}) \boldsymbol{I}. \tag{3.5}$$

The quantity $h = \ln \Sigma + \text{const.}$ is introduced, which transforms the continuity equation into

$$\frac{\partial h}{\partial t} + \boldsymbol{v} \cdot \nabla h + \nabla \cdot \boldsymbol{v} = 0. \tag{3.6}$$

The $\nabla \Phi_{\mathrm{d,m}}$ term is evaluated with the use of the Fourier method in order to solve the Poisson equation

$$\nabla^2 \Phi_d = 4\pi G \Sigma \delta(z), \tag{3.7}$$

where $\delta(z)$ is the Dirac delta function; the resulting expression for the disc potential in Fourier space, where it is most easily expressed, is

$$\widetilde{\Phi}_{\mathrm{d}} = -\frac{2\pi G \widetilde{\Sigma}}{\sqrt{k_x^2 + k_y^2}} \mathrm{e}^{-|\boldsymbol{k}||z|}, \tag{3.8}$$

where G is the gravitational constant, k_x and k_y represent the radial and azimuthal components of the wavevector $\boldsymbol{k} \neq 0$ and z is the height above the disc midplane. From the above, it is easy to obtain the expression for the disc potential at the

midplane ($z = 0$)

$$\widetilde{\Phi}_{\mathrm{d,m}} = -\frac{2\pi G\widetilde{\Sigma}}{\sqrt{k_x^2 + k_y^2}}. \tag{3.9}$$

The (specific) internal energy e of the system must also be considered, its temporal evolution being given by:

$$\frac{\partial e}{\partial t} + \boldsymbol{v}\cdot\nabla e = -\frac{P}{\Sigma}\nabla\cdot\boldsymbol{v} + 2\nu_s\boldsymbol{S}^2 + \nu_b(\nabla\cdot\boldsymbol{v})^2 + \frac{1}{\Sigma}\nabla\cdot(\nu_t\Sigma\nabla e) - \frac{e}{\tau_c}. \tag{3.10}$$

The equation features three types of diffusive effects: bulk viscosity (ν_b), shear viscosity (ν_s) and (horizontal) thermal diffusion (ν_t), as well as a cooling time τ_c.

Also important in the analysis carried out in `CASPER` are the potential vorticity (or vortensity) and the entropy (in this case the dimensionless specific entropy), which are defined as

$$\zeta = \frac{2\Omega + (\nabla\times\boldsymbol{v})_z}{\Sigma}, \tag{3.11}$$

$$s = \frac{1}{\gamma}\ln P - \ln\Sigma, \tag{3.12}$$

where γ is the adiabatic index and the pressure can be expressed in terms of e and Σ using

$$P = (\gamma - 1)\Sigma e. \tag{3.13}$$

As the flow evolves from its background state during the simulation, it is useful to rewrite each quantity as a sum of its basic state value and the departure from it, such that $\boldsymbol{v} = \boldsymbol{v_0} + \boldsymbol{v'}$ (where $\boldsymbol{v_0} = (0, -q\Omega x, 0)^T$ and $\boldsymbol{v'} = (u, v, 0)^T$), $\Sigma = \Sigma_0 + \Sigma'$, etc.

The pressure and stress tensor terms in the Navier-Stokes equation can be rewritten in terms of h and e using Equation 3.13; it is also worth mentioning that the terms $2\boldsymbol{\Omega}\times\boldsymbol{v_0} = 2\Omega^2 qx\boldsymbol{e}_x$ and $-\nabla\Phi$ (i.e. the gradient of the effective potential) cancel each other.

The system can then be described by the following set of equations, which can be used for any value of γ (as long as the pressure terms are adequately altered):

$$\mathrm{D}h' = -\left(\partial_x u' + \partial_y v'\right) \equiv -\Delta, \tag{3.14}$$

$$\mathrm{D}u' - 2\Omega v' = -\partial_x \Phi'_{\mathrm{d,m}} - (\gamma - 1)\left(\partial_x e' + e'\partial_x h'\right) + \nu_s \nabla^2 u' + \\ + \left(\nu_b + \frac{1}{3}\nu_s\right)\partial_x \Delta + T_{xx}\partial_x h' + T_{xy}\partial_y h', \tag{3.15}$$

$$\mathrm{D}v' + (2-q)\Omega u' = -\partial_y \Phi'_{\mathrm{d,m}} - (\gamma - 1)\left(\partial_y e' + e'\partial_y h'\right) + \nu_s \nabla^2 v' + \\ + \left(\nu_b + \frac{1}{3}\nu_s\right)\partial_y \Delta + T_{yx}\partial_x h' + T_{yy}\partial_y h', \tag{3.16}$$

$$\mathrm{D}e' = -(\gamma - 1)e'\Delta + \nu_s S + \left(\nu_b + \frac{2}{3}\nu_s\right)\Delta^2 + \nu_t \nabla^2 e' - \frac{\Omega e'}{\tau_c}, \tag{3.17}$$

with the above equations having been simplified by means of the following quantities:

$$T_{xx} = 2\nu_s \partial_x u' + \left(\nu_b - \frac{2}{3}\nu_s\right)\Delta, \tag{3.18}$$

$$T_{yy} = 2\nu_s \partial_y v' + \left(\nu_b - \frac{2}{3}\nu_s\right)\Delta, \tag{3.19}$$

$$T_{xy} = T_{yx} = \nu_s\left(-q\Omega + \partial_x v' + \partial_y u'\right), \tag{3.20}$$

$$S = 2\left(\partial_x u'\right)^2 + 2\left(\partial_y v'\right)^2 + \left(-q\Omega + \partial_x v' + \partial_y u'\right)^2. \tag{3.21}$$

3.3 Specifications

3.3.1 Diffusive processes

Viscosity

The code makes use of two types of viscosity: bulk and shear. Bulk viscosity only plays an important role in compressive motions, meaning it is often overlooked except in compressible situations like resolving shocks. Shear viscosity, on the other hand, is important because of the differential nature of accretion discs' rotation; it extracts energy from the velocity gradient of the flow and causes a radial (i.e. orthogonal to the flow) transport of angular momentum within the disc (Frank et al., 2002).

For reasons of simplicity the kinematic viscosity coefficients ν_b and ν_s are assumed to be independent of radius, surface density or temperature. This also implies that the flow considered would be viscously stable as

$$\frac{\partial(\nu\Sigma)}{\partial\Sigma} > 0. \tag{3.22}$$

As seen in Equations 3.4– 3.5, these viscosities give rise to a viscous stress $\boldsymbol{T}$ which is defined as

$$\boldsymbol{T} = 2\mu_s \left(\frac{1}{2} \left[\nabla \boldsymbol{v} + (\nabla \boldsymbol{v})^T \right] - \frac{1}{3} (\nabla \cdot \boldsymbol{v}) \boldsymbol{I} \right) + \mu_b (\nabla \cdot \boldsymbol{v}) \boldsymbol{I}, \tag{3.23}$$

which in turn produces a heating rate

$$\begin{aligned} \mathcal{H}_\nu = \boldsymbol{T} : \nabla \boldsymbol{v} \ &= 2\mu_s \left(\frac{1}{2} \left[\nabla \boldsymbol{v} + (\nabla \boldsymbol{v})^T \right] - \frac{1}{3} (\nabla \cdot \boldsymbol{v}) \boldsymbol{I} \right)^2 + \mu_b (\nabla \cdot \boldsymbol{v})^2 \\ &= 2\mu_s \boldsymbol{S}^2 + \mu_b (\nabla \cdot \boldsymbol{v})^2. \end{aligned} \tag{3.24}$$

As mentioned previously the kinematic viscosities are expected to be constant throughout the simulation, their value being set in the initial conditions. If however the chosen resolution is not high enough for strong shocks to be adequately resolved by the method, this can over-rule the constant viscosity value and increase it to avoid numerical errors. Specifically, at every timestep the viscosity ν_i^t is set to the larger of the initial constant viscosity ν_i^0 and that required to resolve shocks according to

$$\nu_i^t = \max \left(\nu_i^0,\, c_i U_{\max} \Delta x,\, c_i V_{\max} \Delta y \right), \tag{3.25}$$

where i denotes a specific viscosity type, c_i is a safety factor (whose value is typically chosen to be $\sim 0.1 - 0.15$), Δx and Δy are the spatial resolutions in the radial and azimuthal directions and $U_{\max}$ and $V_{\max}$ are the maximum wave velocities attained in the $x-$ and $y-$directions, given by

$$U_{\max} = |u|_{\max} + c_s \tag{3.26}$$

$$V_{\max} = |v|_{\max} + c_s. \tag{3.27}$$

It is crucial to enforce this condition in numerical methods as an insufficiently large dissipation can result in numerical instabilities or spurious effects. An example of this was encountered when solving the Burgers' equation as part as one of the tests for the code (Section 3.5.1). A spurious oscillation was observed on top of the velocity profile, which was identified as the Gibbs phenomenon; this is the peculiar behaviour a truncated Fourier series representing a periodic function exhibits near a discontinuity (Hewitt and Hewitt, 1979). In the test run concerned, which featured only a bulk viscosity and where the units implied $c_{\text{iso}} = 1$, the significance of the Gibbs oscillations was found to depend on the parameter $\nu_b/(c_{\text{iso}}\Delta x)$, with the oscillations becoming visible on the solution for $\nu_b/(c_{\text{iso}}\Delta x) \lesssim 0.15$ and overwhelming the solution for $\nu_b/(c_{\text{iso}}\Delta x) \lesssim 0.03$. Figure 3.2 gives an example of the two cases of intensity of the oscillations.

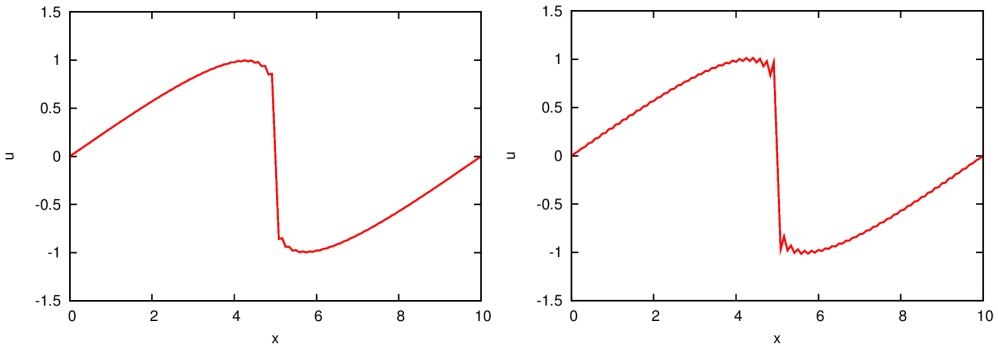


Figure 3.2 – Examples of spurious oscillations due to the Gibbs phenomenon on the solution to Burgers' equation for cases where the oscillations ***left:*** are just visible for $\nu_b/(c_{\text{iso}}\Delta x) = 0.13$ and ***right:*** overwhelm the solution for $\nu_b/(c_{\text{iso}}\Delta x) = 0.018$.

Thermal diffusion

The code, aside from bulk and shear viscosities and a uniform cooling, also takes into account the effect of (horizontal) thermal diffusion; this represents the way heat diffuses from hot regions of the disc to cooler ones, and is defined as

$$\nu_t = \kappa_t/\rho C_p, \tag{3.28}$$

where κ_t is the thermal conductivity coefficient and C_p the fluid's specific heat capacity at constant pressure. The introduction of thermal diffusivity has the

effect of smoothing out the disc's temperature profile, damping any extreme local thermodynamics, and therefore aiding the stability of the disc.

In the `CASPER` code, just like for the bulk and shear viscosities, the thermal diffusion is assumed to be independent of any quantity but the simulation time, with its evolution being dictated by Equation 3.25.

3.3.2 Time stepping

The length of the simulation's timestep was adapted to the ever-changing properties of the flow so as to avoid numerical instability.

If one defines the advection (in $x-$ and $y-$ directions) and viscous timescales as

$$\tau_{\mathrm{adv},x} = C_{\mathrm{CFL}} \frac{\Delta x}{U_{\max}} \tag{3.29}$$

$$\tau_{\mathrm{adv},y} = C_{\mathrm{CFL}} \frac{\Delta y}{V_{\max}} \tag{3.30}$$

$$\tau_{\mathrm{visc}} \simeq C_{\nu} \frac{1}{\nu k_{\max}^2}, \tag{3.31}$$

then the algorithm for determining the timestep is

$$\Delta t = \min\left(\tau_{\mathrm{visc}}\,, \min\left(\tau_{\mathrm{adv},x}, \tau_{\mathrm{adv},y}\right)\right), \tag{3.32}$$

where C_{ν} and C_{CFL} are safety factors, so that the computed timestep is well inside of the viscous (C_{ν}) and advective (C_{CFL}) timescales, the latter being dictated by the Courant–Friedrichs–Lewy – or CFL – condition.

It is worth noting that the viscous timescale employs the general kinematic viscosity ν, whose value is chosen according to the adaptive algorithm

$$\nu = \max(\nu_b + \nu_s\,, \nu_t). \tag{3.33}$$

3.3.3 Shear and remapping

The effect of the disc's differential rotation in the shearing sheet model is taken into account with the shearing of the coordinates in both real and Fourier space;

namely the coordinates being affected are the y direction in real space and the k_x in Fourier space, whose values at a time $t + \Delta t$ obey the following relations:

$$y(t + \Delta t) = y(t) - xq\Omega\Delta t, \tag{3.34}$$

$$k_x(t + \Delta t) = k_x(t) + k_y q\Omega\Delta t. \tag{3.35}$$

It is important to notice that k_x presents a periodicity which is caused by computational shortcomings. While a shearing wave would ideally shear in the direction dictated by the sign of k_y until $|k_x| \to \infty$, this is of course not reproducible in a computational method; instead, the k-space is truncated by adopting a discrete Fourier transform, which makes it periodic. This simplification comes with some caveats: energy is not perfectly conserved and aliasing errors may occur, as described in Section 3.3.4.

The periodic boundary conditions however imply that some remapping must occur when the coordinates shear out of the y or k_x domains considered, as illustrated in Figure 3.3. Here a mode A having $k_x < 0$ shears towards more negative k_x due to finding itself in the negative half of the k_y domain; after a particular time step, its wavenumber is out of bounds of the k_x domain by an amount Δk. Due to the periodic boundary conditions of the problem, it is remapped on the opposite side of the k_x range with its new radial wavenumber being $k_x = k_x^+ - \Delta k$, with k_x^+ being the maximum (positive) resolved wavenumber. Its k_y wavenumber remains unchanged.

A similar remapping occurs when the k_x value shears out of the positive end of the range, and also for the y dimension in real space. In general, the remapping rule can be summarised as

$$k^*_{x,\text{remap}} = \begin{cases} k_x^- + (k_x^* - k_x^+) & \text{if } k_x^* > k_x^+ \\ k_x^+ + (k_x^* - k_x^-) & \text{if } k_x^* < k_x^-, \end{cases} \tag{3.36}$$

where k_x^* and $k^*_{x,\text{remap}}$ are the values of the unresolved wavenumber before and after remapping and k_x^- and k_x^+ are the minimum (i.e. largest negative) and maximum resolved wavenumbers in that same dimension. An equivalent rule also applies to remapping in the y direction in real space.

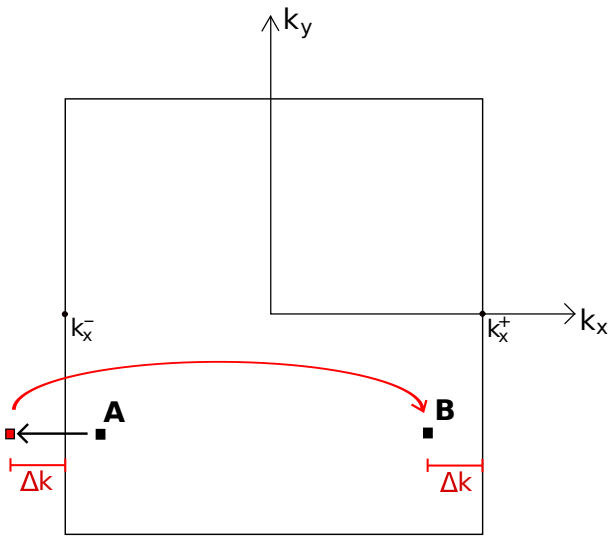


Figure 3.3 – Representation of the remapping due to the periodic boundary conditions in a shearing system. Mode A shears to more negative k_x values due to being in the negative half of the k_y domain and ends up out of the shearing sheet considered by an amount Δk; because of the periodic boundary conditions imposed, the mode gets remapped at point B at the other end of the k_x range, with $k_x = k_x^+ - \Delta k$ and its original k_y value unaltered.

3.3.4 Anti-aliasing

The `CASPER` code also includes an anti-aliasing filter, which can be easily turned on or off by changing the value of a parameter before running the simulation. Aliasing is due to distinct signals becoming indistinguishable from (or aliases of) one another due to insufficient sampling. In the case of waves, this occurs when waves of different wavelengths can both fit the sample available, as shown in Figure 3.4.

In a spectral method aliasing is due to high wavenumber values created by the non-linear terms in the equations of the system. While in an ideal, infinite k domain this does not present an issue, in a discrete Fourier transform domain this causes some of these wavenumber values to exceed the maximum $|k_x|_{\max}$ and $|k_y|_{\max}$ values available. These modes are therefore unresolved, as their wavenumbers' absolute values are larger than the Nyquist wavenumber $k_{\text{Nyquist}} = 2\pi/\lambda_{\text{Nyquist}}$, where $\lambda_{\text{Nyquist}} = 2\Delta x$ is by definition the shortest possible resolvable wavelength. The periodic nature of the truncated k domain means that these unresolved wavenumbers are remapped according to the periodic boundary conditions of the system, as shown in Figure 3.3 and Equation 3.36. This essentially means that in the k_x direction a trailing wave with $k_x^* > k_x^+$ would be spuriously remapped back as a leading

wave and vice versa, with the same process also applying in the k_y direction.

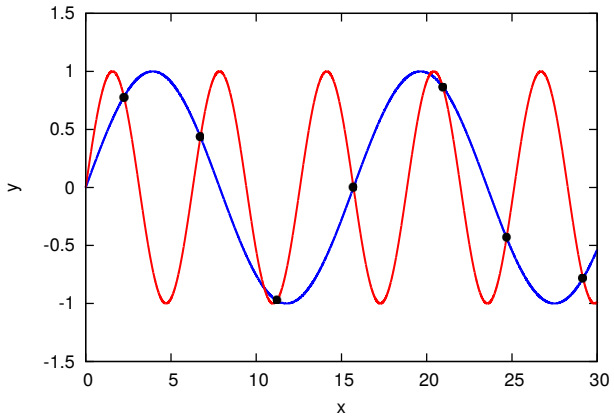


Figure 3.4 – Example of aliasing with two waves of different wavelengths being indistinguishable from one another due to the low sampling rate used

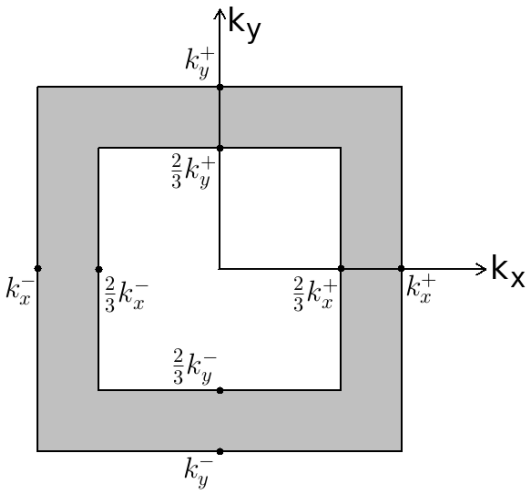


Figure 3.5 – Aliasing mask for the 2/3-rule or truncation method. The values of all the parameters having wavenumber coordinates falling into the shaded area are discarded.

To correct this aliasing error, the 2/3-rule (or truncation) was used as it presented the simplest and computationally least expensive choice. Truncation consists in applying a mask to the chosen box, where the value of any parameter having a wavenumber falling into the last 1/3 frame of the domain (in both k_x and k_y) is set to zero, as shown in Figure 3.5. This means that whenever non-linear terms produced an unresolved k^* value, this would be mapped back into the discarded part of the domain, thus drastically decreasing the likelihood of an aliasing issue

occurring. Of course, this also means that some of the energy in the simulation would be lost as discarding the value falling in the region of interest violates the conservation of energy, but as the truncation is applied at every sub-step within the Runge-Kutta method, the loss would be minimal. It is important to notice that the anti-aliasing correction only works for quadratic non-linear terms, meaning it does not affect gravitational effects.

3.3.5 Gravitational softening

The effects of self-gravity play a fundamental part in the dynamics of the system simulated by `CASPER`. However, the code's 2-dimensional nature means that the disc considered is one where all the points have been projected onto its midplane; this flattening of the disc, which would always have some finite thickness in real life, causes an increase in the gravitational interaction between any two points within the disc. This is because their 3D distance $r_{3\mathrm{D}}$ is reduced to a smaller 2D distance $r_{2\mathrm{D}}$ due to the vertical distance $(z_2 - z_1)$ term dropping, as shown in Equation 3.37

$$\begin{aligned} r_{3\mathrm{D}} &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \\ r_{2\mathrm{D}} &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \end{aligned} \tag{3.37}$$

where (x_1, y_1, z_1) and (x_2, y_2, z_2) are the spatial coordinates of the two points being considered.

This reduction in the two points' separation, which is simply illustrated in Figure 3.6, is most pronounced when the points being considered are closer horizontally than they are vertically.

To correct for the resulting over-estimation of the gravitational interaction between any two fluid elements, a softening length s is introduced so as to modify their distance to

$$r_{2\mathrm{D}}^{\mathrm{soft}} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + s^2}. \tag{3.38}$$

This modifies the midplane disc potential to

$$\widetilde{\Phi}_{\mathrm{d,m}} = -\frac{2\pi G\widetilde{\Sigma}}{\sqrt{k_x^2 + k_y^2}} \mathrm{e}^{-s|k|}, \tag{3.39}$$

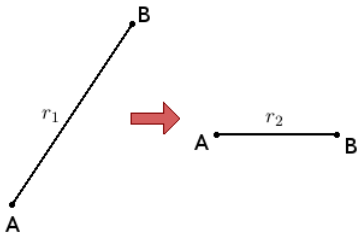


Figure 3.6 – The flattening of the disc from three to two spatial dimensions causes the distance between two points to diminish, unless they are situated at the same z value, hence increasing the gravitational interaction between them.

which also results in the following modification to the dispersion relation for density waves:

$$\omega^2 = \kappa^2 - 2\pi G|k|\mathrm{e}^{-s|k|} + c_s^2 k^2. \tag{3.40}$$

3.4 Data outputs

Owing to the importance of the analysis of the data produced by the code, `CASPER` produces a flexible and vast range of data outputs, in both Fourier and real space.

Being a spectral method, the code allows the analysis of a wide range of quantities in Fourier space both as time profiles and as power spectra to identify the role played by potential instabilities, zonal flows and the interplay between them and (leading and trailing) density waves. The spectral analysis, for instance, makes it possible to quickly and precisely pinpoint the wavelengths of the zonal flows and the non-axisymmetric instability so that their analysis can be deepened. It is also extremely useful in identifying whether the energy of the system is primarily held in compressive or vortical disturbances and the role that instabilities play in the feedback cycle required to sustain turbulence.

The output of real space quantities permits us to look into the temporal development of key quantities such as pressure, density, internal energy and Reynolds and gravitational stresses to pinpoint their roles in the dynamics of the system. It is however also necessary to look at these quantities in the $x - y$ domain in order to observe their spatial structure as well as their temporal evolution. In order to do this, a `VTK` file is produced every n timesteps in order to construct a 2D animation of the chosen quantities. One still image of such an animation is shown in

Figure 3.7 where the entropy of the system is plotted.

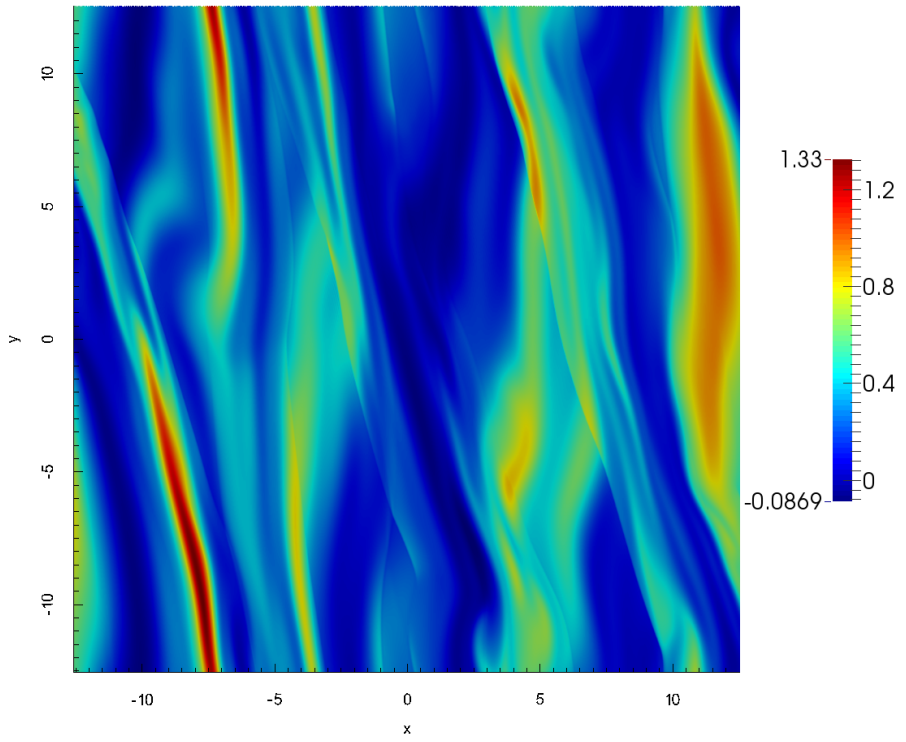


Figure 3.7 – VTK output showing the entropy in (x,y) space

3.5 Computational tests

3.5.1 Solving Burgers' equation

The first test in the development of `CASPER` was to solve Burgers' equation; this is a fundamental 1D PDE named after Dutch physicist Johannes Martinus Burgers, which occurs in numerous areas of applied mathematics. Its viscous form is given by

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu_b \frac{\partial^2 u}{\partial x^2}, \tag{3.41}$$

where in this instance the bulk viscosity ν_b was chosen to be a time-independent constant, $\nu_b \frac{\partial^2 u}{\partial x^2}$ is the diffusive term and $u\frac{\partial u}{\partial x}$ is the non-linear advective term. If the diffusive term is neglected, one obtains the inviscid form of Burgers' equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0. \tag{3.42}$$

The simplicity of Burgers' equation makes it a good candidate for a first benchmark test in the development of any new fluid dynamics code; but it is the presence of diffusive and advective effects that makes it especially suitable to the purpose. Another key aspect of Burgers' equation is the formation of discontinuities, which provided the first test of resolving shocks for the `CASPER` code (the matter is looked at in more detail in Section 3.5.5).

Burgers' equation was successfully solved, with bulk viscosity values of $\nu_b \approx 0.05$ and a fixed timestep of $\Delta t = 10^{-3}$ (in a system of units where the initial velocity $u_0 = 1$ and the box size $L = 10$); simplifying methods such as the integration method and the Cole-Hopf transformation were also tried, with further details and relevant plots found in Appendix A.

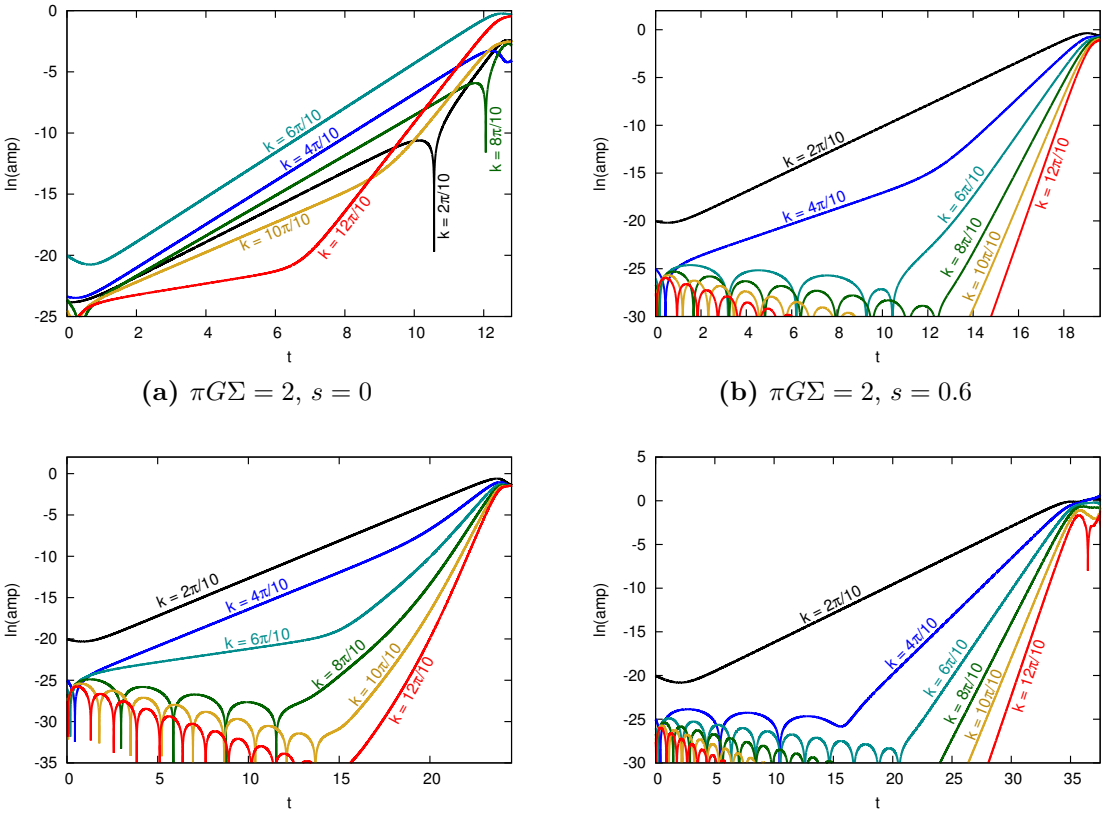
3.5.2 Axisymmetric instability in a non-rotating system

Back to the 2D shearing sheet, a key test of the code was to analyse the Jeans instability triggered when a non-rotating system is perturbed with an axisymmetric disturbance. The disturbance applied to the system coincides with the fastest growing mode of the system (where the units are such that $c_{\text{iso}} = 1$, $L = 10$); this choice has however no effect on the dynamics of the flow. The instability is helped by the temporary absence of the stabilising effect of the Coriolis force at this point in the development of `CASPER`.

The growth resulting from the axisymmetric instability is of an exponential nature, so the temporal development of the key quantities' natural logarithm is analysed; this means that the time evolution of an unstable mode would be a straight line, with its gradient being the growth rate of the instability. Figure 3.8 shows the development of $\ln \widetilde{h}'$ for the six smallest radial wavenumbers $k_i = i\frac{2\pi}{L}$ resolved in the simulation.

The plots show a variety of points worthy of attention:

- The introduction of a softening length – its theoretical background and motivation explained in Section 3.3.5 – stabilises a few modes (four for the case of $\pi G\Sigma = 2$ and two for $\pi G\Sigma = 1$, respectively), highlighting the potential



(a) $\pi G\Sigma = 2,\ s = 0$ **(b)** $\pi G\Sigma = 2,\ s = 0.6$

(c) $\pi G\Sigma = 1,\ s = 0$ **(d)** $\pi G\Sigma = 1,\ s = 0.6$

Figure 3.8 – Natural log of the perturbation amplitude of h' ($\ln \Sigma$) in Fourier space over time for the first six wavenumbers for $\nu_b = 0.1$ and $N = 100$, where the units are such that $c_{\text{iso}} = 1$, $L = 10$. The intrinsic growth rates can be extrapolated by calculating the corresponding line gradients for the appropriate modes. Only certain modes are intrinsically unstable, while some are stable and are seen to simply oscillate while being damped; eventually the fastest growing mode in each case forces stable modes into an instability and previously unstable modes into a faster growth. Exponential growth saturates as the modes enter a non-linear regime.

importance of the increased gravitational interaction due to flattening a 3D disc into a 2D one.

In the case of $\pi G\Sigma = 2$ it also affects the value of the wavenumber of the fastest growing mode, with the $k = 6\pi/10$ mode showing the largest growth rate for $s = 0$ but the $k = 2\pi/10$ one taking over as the fastest growing when $s = 0.6$. This is of course due to the gravitational softening factor e^{-sk} introduced in the self-gravitational potential, which damps higher k modes

more (as seen in Equation 3.40 in the case of the dispersion relation).

- At the beginning of the simulation, exponential growth and decay compete for a short time; eventually exponential growth dominates over the decay and a straight line is obtained for the intrinsically unstable modes. The length of the period during which exponential growth and decay compete appears to be lengthened by the introduction of the softening length.

- Only a fraction of the modes are intrinsically unstable, with the amount of unstable modes being heavily affected by the presence of a softening length. The stable modes are instead seen to oscillate at their respective frequency and decay.

- Intrinsically unstable modes which are not the fastest growing mode are, after some time, forced into a faster growth by non-linear couplings with other modes. For example, in subplot ***(a)***, the mode $k = 10\pi/10$ is a combination of the faster-growing $k = 4\pi/10$ (which grows at a rate $\gamma_{4\pi/10}$) and the $k = 6\pi/10$ mode (whose growth rate is $\gamma_{6\pi/10}$), meaning that it will exhibit a forced growth rate of $\gamma_{10\pi/10} = \gamma_{4\pi/10} + \gamma_{6\pi/10}$. In the same subplot, the mode $k = 12\pi/10$ is a combination of the mode $k = 6\pi/10$ with itself, and it is therefore forced to growth at a rate $\gamma_{12\pi/10} = 2\,\gamma_{6\pi/10}$.

  Modes that are intrinsically stable are also forced according to the same non-linear coupling.

- The paths traced by the modes in the above plot are at times seen to exhibit dips (mostly in subplot ***(a)***); this is due to a sign reversal in the wave's amplitude.

- There appears to be a saturation mechanism which halts the progress of the exponential growth, with a non-linear regime following.

In order to check the accuracy of the simulation, it was necessary to compare the obtained growth rates with their theoretical values. To do so, the expected growth rates for each of the four cases considered were calculated using the dispersion relation (with no shear)

$$\omega^2 = c_{\text{iso}} k^2 - 2\pi G \Sigma_0 |k| \mathrm{e}^{-sk}, \tag{3.43}$$

where ω is the wave frequency (where an instability gives $\omega^2 < 0$) and $k = k_x$ due to the axisymmetric nature of the problem.

However, the presence of the bulk viscosity would affect the growth rates so that the observed values would be somewhat smaller (ie, damped) than the ones originating from Equation 3.43. A correction was therefore derived by considering the equations for $\widetilde{u}'$ and $\widetilde{h}'$ ($\widetilde{v}'$ was neglected due to the absence of rotation in the problem)

$$\partial_t \widetilde{u}' = \lambda \widetilde{u}' = -\mathrm{i}k_x \left(\widetilde{h}' + \widetilde{\Phi}_{\mathrm{d,m}} \right) - \nu_b k_x^2 \widetilde{u}', \tag{3.44}$$

$$\partial_t \widetilde{h}' = \lambda \widetilde{h}' = -\mathrm{i}c_{\mathrm{iso}}^2 k_x \widetilde{u}', \tag{3.45}$$

where λ is the growth rate of the solutions assuming an exponential growth of the type $\mathrm{e}^{\lambda t}$. By multiplying the first equation by a factor of λ it is possible to substitute Equation 3.45 into Equation 3.44 to give – after some manipulation – the following expression

$$\lambda \left(\lambda + \nu_b k_x^2 \right) = -c_{\mathrm{iso}}^2 k_x^2 + 2\pi G \Sigma k_x \mathrm{e}^{-sk_x}. \tag{3.46}$$

This strongly resembles the original form of the dispersion relation (Equation 3.43), with $\lambda_0 = \sqrt{\lambda \left(\lambda + \nu_b k_x^2 \right)}$ being the *undamped* growth rate. By completing the square it is possible to obtain λ_0 in terms of the *damped* growth rate λ.

$$\left(\lambda + \frac{1}{2} \nu_b k_x^2 \right)^2 = \lambda_0^2 + \left(\frac{1}{2} \nu_b k_x^2 \right)^2, \tag{3.47}$$

therefore giving the expression to compute the corrected growth rate λ:

$$\lambda = \sqrt{\lambda_0^2 + \left(\frac{1}{2} \nu_b k^2 \right)^2} - \frac{1}{2} \nu_b k^2. \tag{3.48}$$

The first thing to notice about this expression is that – as long as $\lambda_0 > 0$ (ie, there is an instability in the first place) – the corrected value λ is always positive. The viscous damping therefore can substantially weaken the instability for a specific mode, but not remove it altogether. Furthermore, the viscous correction is smaller (by a factor of 2) than the usual $\mathrm{e}^{-\nu k^2}$. This is due to the absence of rotation, which means that the equation for $\widetilde{v}'$ takes no part in the dissipation. It is also worth noting that this correction only directly applies to intrinsic growth rates.

The theoretical values for the intrinsic growth rates are listed in Table 3.2 together with the values $\lambda_{\rm sim}$ obtained from the simulation. It is clear from the

Table 3.2 – Comparison between theoretical and observed intrinsic growth rates for $\nu_b = 0.1$ and $\Delta t = 0.002$. λ_0 and λ represent the growth rate calculated from the dispersion relation for a particular wavelength and its corrected value taking into account viscous dissipation, whilst $\lambda_{\rm sim}$ is the growth rate obtained from the simulation.

Parameters	k	λ_0	λ	$\lambda_{\rm sim}$
$\pi G\Sigma = 2,\ s = 0$	$2\pi/10$	1.4555	1.4360	1.4358490 ± 0.0000008
	$4\pi/10$	1.8567	1.7794	1.7794293 ± 0.0000002
	$6\pi/10$	1.9967	1.8269	$1.82692059 \pm 0.00000009$
	$8\pi/10$	1.9330	1.6428	1.6428567 ± 0.0000006
	$10\pi/10$	1.6422	1.2212	1.221342 ± 0.000002
	$12\pi/10$	0.9313	0.4609	0.46206 ± 0.00001
$\pi G\Sigma = 2,\ s = 0.6$	$2\pi/10$	1.1529	1.1333	1.1332995 ± 0.0000002
	$4\pi/10$	0.8865	0.8110	0.811469 ± 0.000005
$\pi G\Sigma = 1,\ s = 0$	$2\pi/10$	0.9284	0.9088	$0.90882922 \pm 0.00000002$
	$4\pi/10$	0.9666	0.8908	0.890912 ± 0.000002
	$6\pi/10$	0.4657	0.3208	0.321639 ± 0.000008
$\pi G\Sigma = 1,\ s = 0.6$	$2\pi/10$	0.6835	0.6640	0.6640336 ± 0.0000002

tabulated data that the simulation results and the theoretical values match extremely well, with only two values exhibiting a difference greater than 0.2%. The agreement is further enhanced by the small errors observed.

3.5.3 Comparison with linear analysis

Once the Coriolis effect was added into the system, with the equations achieving their full form, a comparison was made with the non-axisymmetric linear analysis whose results are presented in Chapter 4. The results from the linear analysis presented in the next chapter are however obtained for an isothermal system in ideal conditions (no cooling or viscous/diffusive effects), which is not reproducible in `CASPER` as the code requires some viscosity to be numerically stable. A direct comparison was then achieved by adding diffusive effects (bulk and shear viscosities only) in the linear analysis, although these are not used in Chapter 4.

A sinusoidal axisymmetric structure of amplitude A_h and wavenumber k is therefore imposed on the ICs in `CASPER`, along with a non-axisymmetric perturba-

tion possessing a wavenumber k_y; the growth rate is calculated for a range of Q values, with the results of the comparison being shown in Figure 3.9.

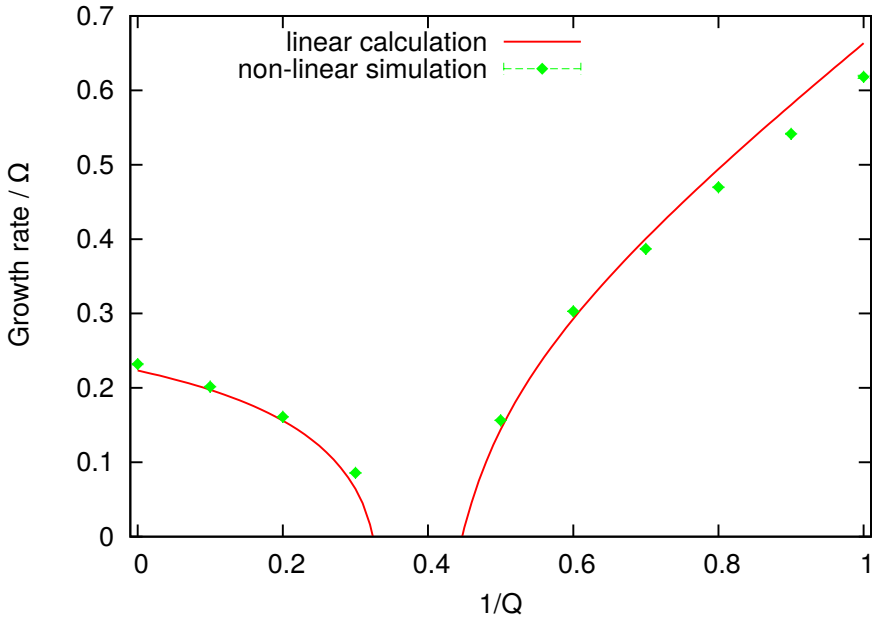


Figure 3.9 – Growth rate comparison between the non-axisymmetric linear analysis (red line) and the `CASPER` code (green points) as a function of Q^{-1} with an imposed axisymmetric structure of amplitude $A_h = 0.1$ and wavenumber $kc_s/\Omega = 5\frac{2\pi}{10}$ and a non-axisymmetric perturbation of wavenumber $k_y c_s/\Omega = \frac{2\pi}{20}$.

The growth rates obtained with the two means of analysis appear to agree with each other in the whole Q^{-1} range. A slight discrepancy is only observed for strong self-gravitating conditions, with the linear analysis recording a somewhat larger growth rate; the discrepancy however remains below 10% at all times. The slight disagreement might be due to the energy loss resulting from the truncation of the Fourier domain, or from the lack of cooling to balance the heating produced by viscous effects; this would in fact be at its strongest for $Q^{-1} \to 1$ and would cause the effective Q^{-1} to decrease, lowering the observed growth rate. While this effect can in principle occur in both the linear analysis and in `CASPER`, the latter is evolved for a much longer time, making the resulting heating of the disc more pronounced.

3.5.4 Shearing wave evolution

Another test carried during the development of `CASPER` is the temporal evolution of individual shearing waves. Their excitation has been suggested as a potential

explanation for various processes occurring in accretion discs, such as increased transport of angular momentum in disc regions where magnetic phenomena are not active (Heinemann and Papaloizou, 2009a).

While the azimuthal wavelength of density waves remains fixed through their temporal evolution, their radial wavenumber is seen to evolve according to

$$k_x(t) = k_x(0) + q\Omega k_y t,$$

where $k_x(0)$ represents the wavenumber value at time $t = 0$. Density wave excitation occurs whenever its radial wavenumber 'swings' through $k_x(t) = 0$, therefore causing the leading density wave to become trailing, instead. Different shearing waves however shear through the wavenumber space at different rates (proportional to their azimuthal wavenumber k_y) and therefore 'swing' from leading to trailing at different times.

Figure 3.10 illustrates the evolution of a single density wave with $k_y = 2\pi/L_y$ from the moment it 'swings' (i.e. starting at $k_x(t = 0) = 0$) to its trailing phase, in both linear (black, dashed line) and non-linear (red, full) regimes, in the radial velocity u and log of density h in Fourier space. The average sound speed of the flow is also shown in the bottom panel. The top and middle panels show the (normalised) imaginary and real parts of the Fourier transforms $\tilde{u}$ and $\tilde{h}$, respectively. In the linear regime the shearing wave is seen to grow after its 'swing', as its frequency gradually increases; when in the non-linear regime, the shearing wave is however gradually damped. This damping, which is too large to be caused by viscous effects, is produced by the formation of weak shocks. As these are dissipated they produce heat, which is recorded in the average sound speed of the disc, as $c_s \propto \sqrt{T}$, by a departure from its initial value.

The shearing wave evolution recorded with `CASPER` matches up well with that recorded by Heinemann and Papaloizou (2009b).

3.5.5 Resolving shocks

The ability to properly capture and resolve the formation and evolution of shocks is not one of the strengths of spectral codes. As seen in Section 3.3.1, an insufficient amount of viscosity in the simulation can result in the development of spurious oscillations (Gibbs phenomenon) on the solution or – in more extreme cases – in numerical errors.

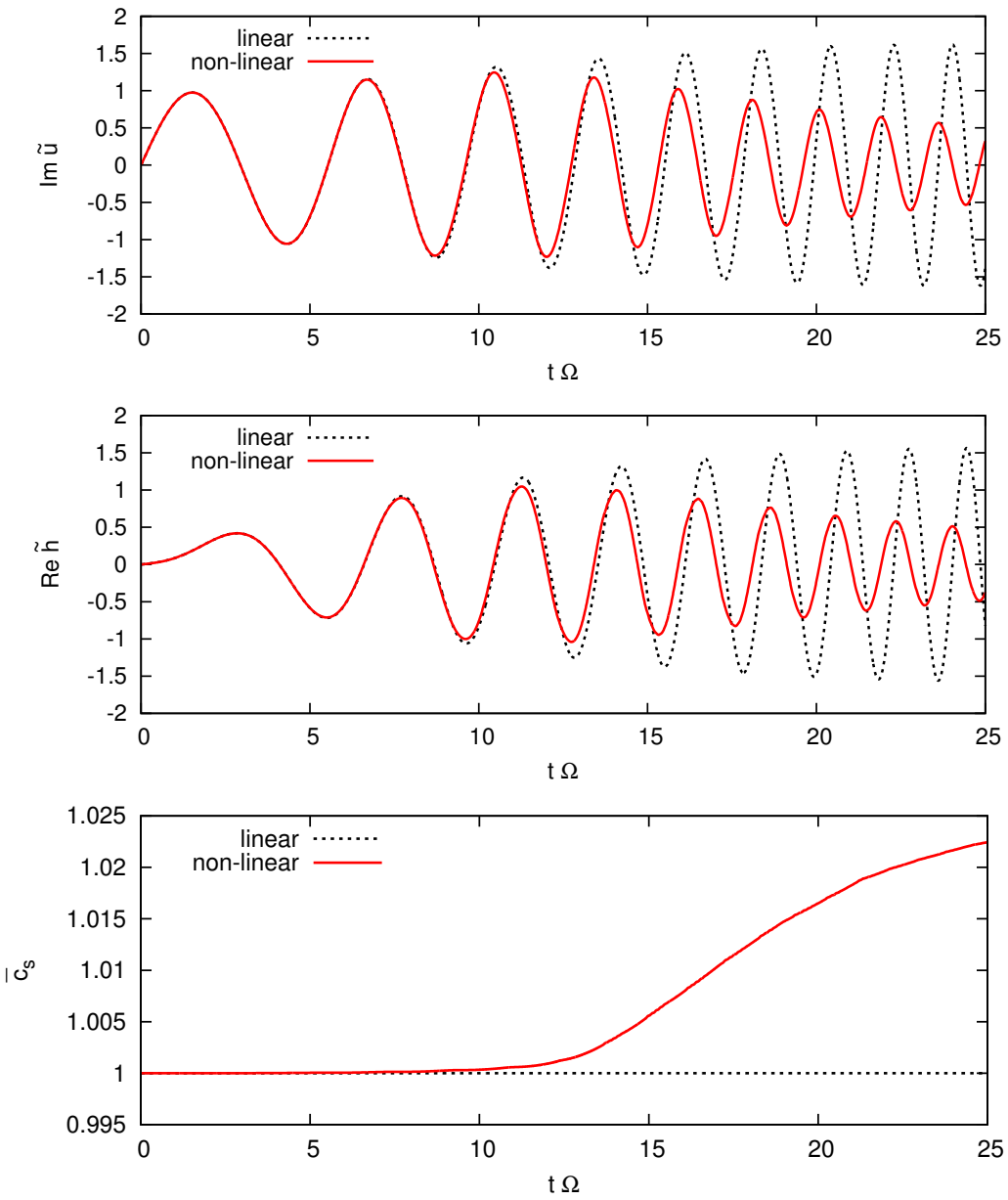


Figure 3.10 – Evolution of a single shearing wave in both linear and non-linear regimes in the (normalised) Fourier space radial velocity (**top**) and h (**middle**), as well as the average sound speed of the flow (**bottom**). The non-linear shearing wave is quickly and gradually damped by weak shocks; the dissipation of such shocks produces heat which acts to increase the sound speed.

It was therefore paramount to test whether the `CASPER` code was able to handle the shocks produced by the system. A first test was carried out by successfully solving Burgers' equation. A more advanced and thorough test was to observe the

x–profile of the 2D shearing sheet radial velocity u at some time t and compare its behaviour when the initial conditions (IC) possessed small and large amplitudes. This is shown in Figure 3.11; the simulation is initialised with small (red, full

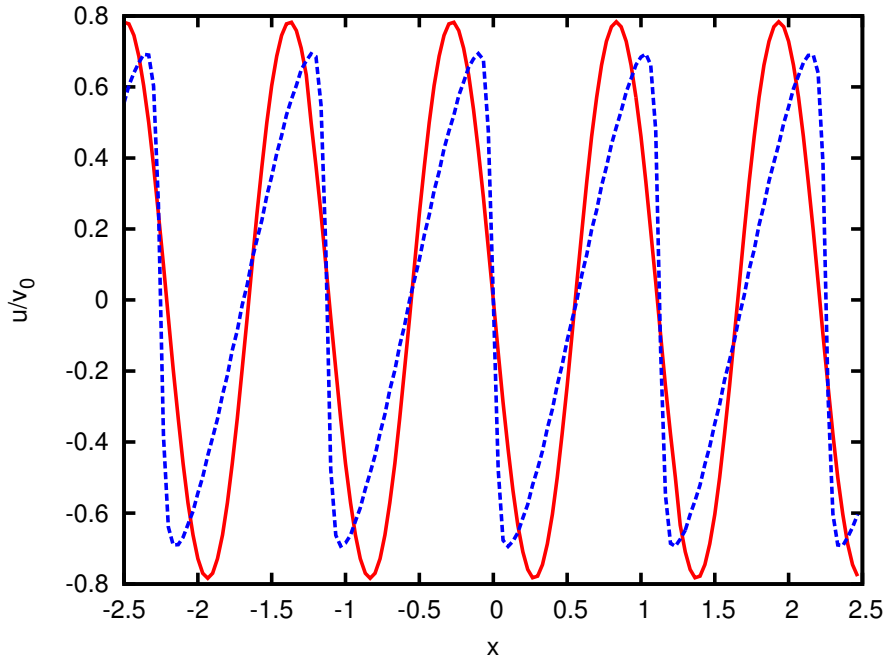


Figure 3.11 – Real space normalised profile of the radial velocity u at time $t = 2.69\Omega^{-1}$ when the applied ICs were of small (red, full) and large (blue, dashed) amplitude, with $\nu_b(\pi G\Sigma_0)^2\Omega^{-3} = 0.003$ and $\nu_s(\pi G\Sigma_0)^2\Omega^{-3} = 0.001$, no gravity and no cooling acting on the system. The profile shows some definite advection in the large amplitude case, with no sign of spurious oscillations or unwanted effects associated with it.

line) and large (blue, dashed) amplitude initial conditions (IC) – with cooling and gravitational effects neglected for the purpose of focusing on the wave dynamics – and the x–profile of u is compared at $t = 2.69\Omega^{-1}$. The large amplitude case clearly shows a substantial amount of advection taking place on the solution, but its profile appears free of the spurious oscillations that can characterise the resolving of shocks in spectral methods. This is regardless of the modest bulk viscosity value $\nu_b(\pi G\Sigma_0)^2\Omega^{-3} = 0.003$ employed. This means `CASPER`'s spectral method is able to properly resolve the shocks forming in the large amplitude case.

A yet more thorough test was to have a quantitative assessment of the code's ability to resolve shocks. Also, it was important to reconsider the problem with cooling and gravitational effects acting on the flow, so as to evaluate the code

in its full form, including the timestepping algorithm and the adaptive viscosity method (described in Equation 3.25) to ensure numerical stability. The code was initialised with white noise in the velocity components but with constant values in h and e, as explained more thoroughly in Section 6.2. As the code's purpose is not that of shock capturing, it was decided to go down the simple yet effective route of locating shocks by means of pressure contours. This enabled to approximately locate shocks, therefore allowing to calculate their Mach number. It was chosen to analyse contours in the pressure – rather than in density, temperature or Mach number (all of which show a bunching up of contour lines in the locus of a shock) – as this does not present a higher concentration of contour lines at the location of a contact discontinuity, unlike the other quantities.

The pressure was therefore obtained in the x–y space by means of `VTK` files, as explained in Section 3.4, with isobar lines being plotted as shown in Figure 3.12; As seen above, the bunching up of pressure contour lines can only arise from shock-

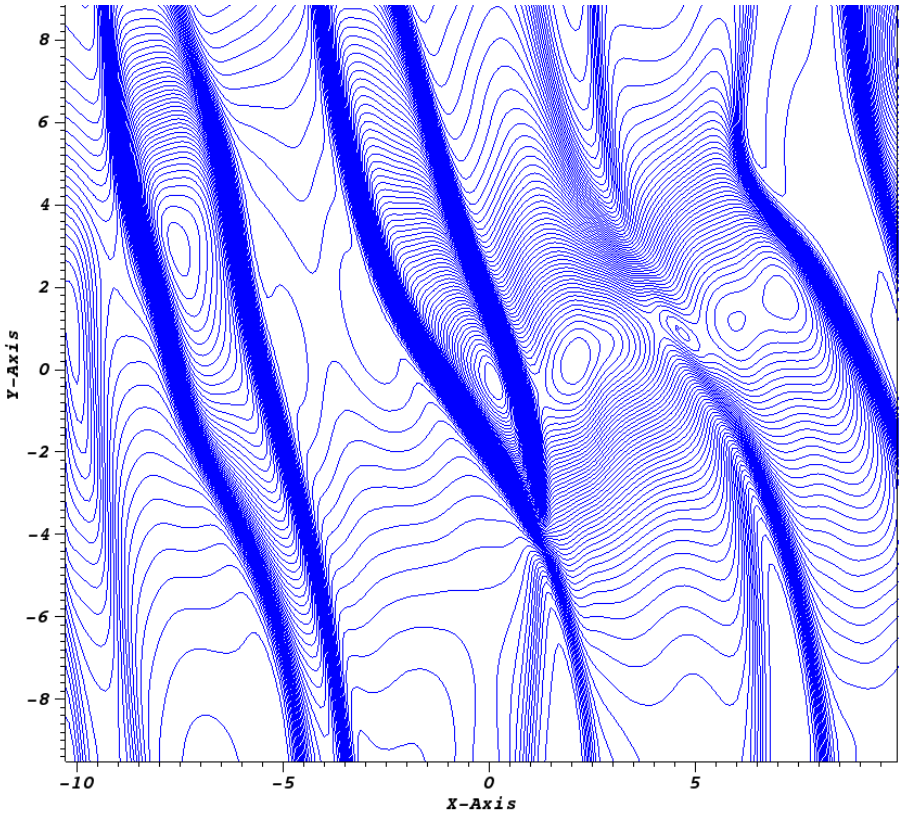


Figure 3.12 – Isobar lines indicating the locations of shock discontinuities where a higher concentration occurs.

driven discontinuities, so these isobars allow to roughly locate shocks. Of course there is no guideline for how bunched up pressure contour lines need to become to be positively identified as the location of a shock, so to largely avoid identifying 'false' shocks, only the few regions with the highest concentration of isobars per frame were analysed. In order to assess whether the code is capturing these shocks appropriately, one can obtain the pressure and density values at either end of a discontinuity and check they agree the Rankine–Hugoniot conditions for shocks. These can be written so as to express the pressure and density ratios across the shock as a function of the Mach number $\mathcal{M}$

$$\frac{P_2}{P_1} = \frac{2\gamma\mathcal{M}^2 - (\gamma - 1)}{\gamma + 1}, \tag{3.49}$$

$$\frac{\rho_2}{\rho_1} = \frac{(\gamma + 1)\mathcal{M}^2}{(\gamma - 1)\mathcal{M}^2 + 2}, \tag{3.50}$$

where the latter equation can be transformed into a difference in h by taking the natural log of both sides and obtaining

$$\Delta h = h_2 - h_1 = \ln\left(\frac{(\gamma + 1)\mathcal{M}^2}{(\gamma - 1)\mathcal{M}^2 + 2}\right). \tag{3.51}$$

Both expressions can therefore be reversed so as to give the Mach number from either pressure ratio or h difference

$$\mathcal{M}_P^2 = \frac{\frac{P_2}{P_1}(\gamma + 1) - (\gamma - 1)}{2\gamma} \tag{3.52}$$

$$\mathcal{M}_h^2 = \frac{2\mathrm{e}^{\Delta h}}{(\gamma + 1) - (\gamma - 1)\mathrm{e}^{\Delta h}}. \tag{3.53}$$

It is apparent from these last two expressions that the values of P and h at either end of a shock can be used to determine two values of the Mach number ($\mathcal{M}_P$ and $\mathcal{M}_h$, respectively) which, if the code can adequately resolve shocks, should be consistent with each other (or equal, in an ideal scenario).

The results of this analysis are shown in Figure 3.13, which shows the relationship between $\mathcal{M}_P^2$ and $\mathcal{M}_h^2$. The dashed line represents the ideal scenario of $\mathcal{M}_P^2 = \mathcal{M}_h^2$, which offers a reference to the shock resolving accuracy of the code. Moreover, the shaded region represents the range of $\mathcal{M}_h^2$ which are within 10% of

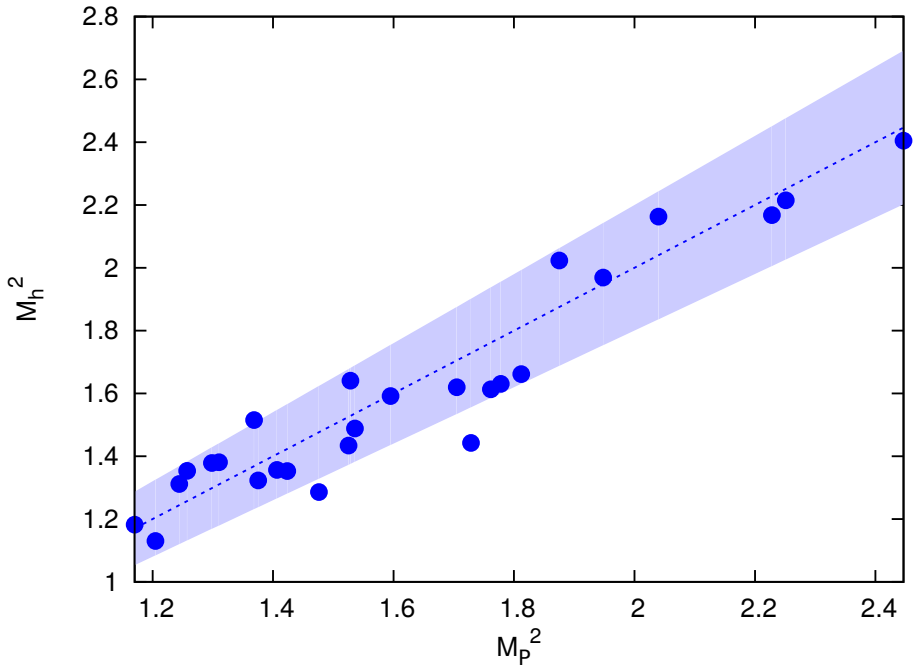


Figure 3.13 – Comparison between the square of Mach number values obtained from the pressure ($\mathcal{M}_P^2$) and density ratios (or h difference, $\mathcal{M}_h^2$) for $\gamma = 2$. The dashed line represents the ideal correlation of $\mathcal{M}_P^2 = \mathcal{M}_h^2$, while the shaded region shows the range of $\mathcal{M}_h^2$ values which are within 10% of the ideal $\mathcal{M}_h = \mathcal{M}_P$ case.

the obtained value for $\mathcal{M}_P^2$; equally, it would have been possible to consider the 10% region of the $\mathcal{M}_h^2$ values, instead. What is clear from this plot is that the code's efficiency at resolving shocks seems to be sufficient for the problem; only three data points lie outside of the 10% region, meaning around 90% of the values lie within it. Furthermore, the accuracy does not appear to decrease as the shocks get stronger, which would have been the case had the code been unable to resolve shocks properly.

It is also worth taking into account that the method of detecting shocks by locating regions of increased isobar line concentration does come with some uncertainty; it is in fact difficult to tell exactly where a shock begins and ends, leading to some uncertainty in the calculated Mach number values. However, this method provides a quick, comprehensive way of establishing that the `CASPER` code is properly able to resolve shocks. The results obtained give confidence that the spectral nature of the code is able to handle shocks without the need to implement a hyper-diffusivity.

Chapter 4

Non-axisymmetric instabilities in discs with imposed zonal flows

4.1 Introduction

The work presented in this chapter represents a stability analysis of an ideal (no viscous or thermal effects) Keplerian disc to non-axisymmetric instabilities when a zonal flow is imposed on the system. As discussed in Section 2.3, zonal flows are observed in a variety of astrophysical settings – both observational and numerical – and in 2D shearing sheet models they are found to break down into vortices due to the action of the Kelvin-Helmholtz instability (Lithwick, 2007; Lithwick, 2009). The analysis carried out by Lithwick (2007) is however of an incompressible nature and further simplifies the problem by only considering azimuthally elongated perturbations such that $k_y \ll k$ (k being the zonal flow wavenumber); this thesis considers instead an extension to such a model, where the flow is treated as fully compressible and no further assumption is made regarding the nature of the disturbances. Furthermore, self-gravity, whose effect is ignored by Lithwick (2007) and Lithwick (2009), is also considered here. The analysis is carried out in both isothermal and adiabatic conditions, with the results appearing mostly insensitive to the change in the adiabatic index value.

The work[1] firstly assesses the suitability of the model by comparing its results directly to those obtained by Lithwick (2007) in a regime where Lithwick (2007)'s analysis is valid. Once the suitability of the analysis has been proven, the work is

[1]The content of this chapter is based on Vanon and Ogilvie (2016)

extended to consider a more general set of conditions and study how the properties of the imposed zonal flow affect the onset and the strength of the Kelvin-Helmholtz instability. Self-gravity is subsequently introduced into the system, with a new instability being observed and the natures of the two instabilities are studied. Lastly the analysis looks into the stability of the system when the zonal flow is present in the entropy, but not in the potential vorticity as in the previous case; an instability of self-gravitating nature is observed, and this appears to shift to weaker self-gravity operating conditions as the strength of the zonal flow is increased; instability is eventually observed for non-SG conditions, possibly confirming the results by Lovelace et al. (1999) although further analysis of this instability is beyond the scope of this work.

4.2 Model

The linear analysis is carried out in compressible conditions, with the simplification of any source of heating (and therefore viscous effects) or cooling being ignored. The inviscid assumption significantly simplifies the set of non-linear equations governing the temporal evolutions of the chosen quantities (the general form of which is given in Equations 3.14–3.17) to the following

$$\mathrm{D}h' = -\left(\partial_x u' + \partial_y v'\right) \equiv -\Delta, \tag{4.1}$$

$$\mathrm{D}u' - 2\Omega v' = -\partial_x \Phi'_{\mathrm{d,m}} - \frac{c_s^2}{\gamma}\partial_x h' - (\gamma - 1)\left[\partial_x e' + e'\partial_x h'\right], \tag{4.2}$$

$$\mathrm{D}v' + (2-q)\Omega u' = -\partial_y \Phi'_{\mathrm{d,m}} - \frac{c_s^2}{\gamma}\partial_y h' - (\gamma - 1)\left[\partial_y e' + e'\partial_y h'\right], \tag{4.3}$$

$$\mathrm{D}e' = -(\gamma - 1)(e_0 + e')\Delta. \tag{4.4}$$

In a linear analysis, the equations above have solutions in the form of shearing waves (Kelvin, 1887) described by

$$\Sigma(\boldsymbol{k}, \boldsymbol{x}) = \Re[\widetilde{\Sigma}'(t)\exp[\mathrm{i}\boldsymbol{k}(t)\cdot\boldsymbol{x}], \tag{4.5}$$

where the tilde ($\tilde{}$) represents the Fourier transform of the original real space quantity. The nature of this solution means that the temporal dependence of $\boldsymbol{k} = (q\Omega k_y t + \text{const.}, \text{const.}, \text{const.})^T$, due to the background shear, cancels the

shear terms $-q\Omega x\frac{\partial h'}{\partial y}$, $-q\Omega x\frac{\partial u'}{\partial y}$ and $-q\Omega x\frac{\partial v'}{\partial y}$ (which are obtained from the $\boldsymbol{v}_0 \cdot \nabla h'$ and $\boldsymbol{v}_0 \cdot \nabla v'$ terms), as shown below

$$\begin{aligned}(\partial_t - q\Omega x\partial_y)u' &= \Re\left\{\left[\partial_t\tilde{u}' + \mathrm{i}\left(\frac{\partial \boldsymbol{k}}{\partial t}\cdot\boldsymbol{x} - q\Omega x k_y\right)\tilde{u}'\right]\mathrm{e}^{\mathrm{i}\boldsymbol{k}(t)\cdot\boldsymbol{r}}\right\} \\ &= \Re\left\{\left[\partial_t\tilde{u}' + \mathrm{i}\left(q\Omega x k_y - q\Omega x k_y\right)\tilde{u}'\right]\mathrm{e}^{\mathrm{i}\boldsymbol{k}(t)\cdot\boldsymbol{r}}\right\} \\ &= \Re\left\{\partial_t\tilde{u}'\mathrm{e}^{\mathrm{i}\boldsymbol{k}(t)\cdot\boldsymbol{r}}\right\}. \end{aligned} \tag{4.6}$$

Alongside these non-axisymmetric solutions, which are either density waves or non wave-like disturbances of entropy or potential vorticity, there also exist axisymmetric ones (i.e. possessing $k_y = 0$, $k_x = \text{const.}$) which represent density wave or stationary structures in the same quantities.

4.2.1 The axisymmetric structure

The axisymmetric structure being imposed on the system is a zonal flow possessing a sinusoidal profile and being accompanied by density (and hence pressure) variations in order to maintain the geostrophic balance. Its form is given by

$$\begin{aligned} h'_{\mathrm{str}} &= A_h\cos(kx), \\ v'_{\mathrm{str}} &= A_v\sin(kx), \\ e'_{\mathrm{str}} &= e_0 A_e\cos(kx), \end{aligned} \tag{4.7}$$

where A_h, A_v and A_e represent the amplitudes of the axisymmetric structure in the respective quantities, and k is the structure's wavenumber. The analysis is conducted to first order in the amplitudes.

The expression for the surface density disturbance Σ' can be found, using the equation $\Sigma_0 + \Sigma' = \Sigma_0 e^h$, from h'_{str} by doing a Taylor series in A_h. This results in

$$\frac{\Sigma'}{\Sigma_0} = A_h\cos(kx) + \mathcal{O}(A_h^2). \tag{4.8}$$

It is then finally possible to obtain an expression for the disc potential perturbation using Equation 3.8

$$\Phi'_{\mathrm{d,m}} \approx -\frac{2\pi G\Sigma_0}{\sqrt{k_x^2+k_y^2}} A_h \cos(kx). \tag{4.9}$$

The equation of geostrophic balance can then be written in terms of the amplitudes

$$-2\Omega A_v = -2\pi G\Sigma_0 A_h + \frac{c_s^2}{\gamma} k(A_e + A_h). \tag{4.10}$$

A significant detail to be noticed is that in the isothermal scenario (for which the value of the adiabatic index is set to $\gamma = 1$ and any variations in the internal energy are neglected, causing $A_e = 0$) with no SG, the equations describing the system would be exact. This is because the disc self-gravitational potential shown above and the non-linear pressure term of Equation 4.2 would be discarded.

An expression for A_v can then be deduced from Equation 4.10

$$A_v = \frac{c_s \kappa}{Q\Omega} A_h - \frac{c_s^2 k}{2\gamma\Omega}(A_e + A_h), \tag{4.11}$$

where Q and κ are the Toomre's parameter characterising the disc's stability to self-gravity and the epicyclic frequency, as previously mentioned in Section 2.2.2.

The presence of an axisymmetric structure in h, v and e also causes a similar structure to emerge in the potential vorticity and specific entropy; the amplitudes of the disturbances in these two quantities can be expressed in terms of the other amplitudes, making use of the equation of geostrophic balance, as follows

$$\begin{aligned} s' &= \frac{P'}{\gamma P} - \frac{\Sigma'}{\Sigma} = \frac{1}{\gamma}\left(\frac{\Sigma'}{\Sigma} + \frac{e'}{e}\right) - \frac{\Sigma'}{\Sigma} \\ &= \frac{1}{\gamma}\left(A_e - (\gamma - 1)A_h\right)\cos(kx) \\ &= A_s \cos(kx), \end{aligned} \tag{4.12}$$

$$\begin{aligned}\frac{\zeta'}{\zeta} &= \frac{\partial_x v'}{(2-q)\Omega} - \frac{\Sigma'}{\Sigma} \\ &= -\frac{1}{\kappa^2}\left[\left(A_h + A_s\right) c_s^2 k^2 - A_h 2\pi G\Sigma k\right]\cos(kx) - A_h\cos(kx) \\ &= -\frac{1}{\kappa^2}\left[\left(c_s^2 k^2 - 2\pi G\Sigma k + \kappa^2\right) A_h + c_s^2 k^2 A_s\right]\cos(kx) \\ &= A_\zeta \cos(kx). \end{aligned} \tag{4.13}$$

Given the expressions for A_s and A_ζ – as well as the dependence of A_v on A_h and A_e – it is possible to fully describe the imposed structure by means of only two parameters: A_e and A_h, or A_s and A_ζ.

The emergence of an axisymmetric structure in the potential vorticity can result in the system being unstable to axisymmetric disturbances according to the Rayleigh criterion. This usually states that for the system to be unstable to axisymmetric perturbations, the radial gradient of its specific angular momentum j must be negative

$$\frac{\mathrm{d}j^2}{\mathrm{d}r} < 0. \tag{4.14}$$

In the rotating frame of reference of the shearing sheet model used in this analysis, this occurs when the vorticity due to the presence of the zonal flow exceeds the value of the background vorticity (due do the former being negative)

$$2\Omega\left((2-q)\Omega + \frac{\mathrm{d}v}{\mathrm{d}x}\right) < 0. \tag{4.15}$$

For the axisymmetric structure considered in this analysis in non-SG conditions, this occurs when

$$A_h\left(\frac{kc_s}{\Omega}\right)^2 > 2(2-q), \tag{4.16}$$

with $2(2-q) = 1$ for a Keplerian disc. Our analysis is however unable to pick up this local axisymmetric instability which in any case requires a third dimension to operate. Nevertheless, the above version of the Rayleigh criterion is taken into consideration when analysing the results of this work, as it could provide a natural limitation to the lengthscale of the zonal flows involved.

4.2.2 Introducing non-axisymmetric perturbations

When non-axisymmetric disturbances are applied to the described system, a ladder of shearing waves is generated. This shearing wave ladder, which is shown in Figure 4.1, takes the form

$$u_i' = \Re \sum_{n=-\infty}^{\infty} \{\tilde{u}_n'(t) \exp\left[\mathrm{i}k_{xn}(t)x + \mathrm{i}k_y y\right]\}, \tag{4.17}$$

with similar expressions for v_i', h_i' and e_i', where

$$k_{xn}(t) = k_{x0}(0) + nk + q\Omega k_y t \tag{4.18}$$

is the wavenumber in the radial direction of the shearing ladder modes, with k_{x0} being the value for mode n of the ladder at $t = 0$ and $n \in [-N, N]$.

The presence of the imposed axisymmetric structure means that each mode n of the shearing wave ladder is coupled to its neighbours $n-1$ and $n+1$, as shown by the equations governing the system given below

$$\partial_t \tilde{h}_n' = \frac{kA_h}{2\mathrm{i}} \left(\tilde{u}_{n-1}' - \tilde{u}_{n+1}'\right) - \frac{A_v k_y}{2} \left(\tilde{h}_{n-1}' - \tilde{h}_{n+1}'\right) - \mathrm{i}\left(k_{xn}\tilde{u}_n' + k_y \tilde{v}_n'\right) \tag{4.19}$$

$$\begin{aligned}\partial_t \tilde{u}_n' = &-\frac{A_v k_y}{2} \left(\tilde{u}_{n-1}' - \tilde{u}_{n+1}'\right) + 2\Omega\tilde{v}_n' - \mathrm{i}k_{xn}\left(\frac{v_s^2}{\gamma} - 2\frac{v_s\kappa}{|k_n|Q}\right)\tilde{h}_n' \\ &- (\gamma - 1)\left[\frac{\mathrm{i}e_0 A_e}{2}\left(k_{x,n-1}\tilde{h}_{n-1}' + k_{x,n+1}\tilde{h}_{n+1}'\right) - \frac{kA_h}{2\mathrm{i}}\left(\tilde{e}_{n-1}' - \tilde{e}_{n+1}'\right) + \mathrm{i}k_{xn}\tilde{e}_n'\right]\end{aligned} \tag{4.20}$$

$$\begin{aligned}\partial_t \tilde{v}_n' = &-\frac{A_v}{2}\left[k\left(\tilde{u}_{n-1}' + \tilde{u}_{n+1}'\right) + k_y\left(\tilde{v}_{n-1}' - \tilde{v}_{n+1}'\right)\right] - (2-q)\Omega\tilde{u}_n' - \mathrm{i}k_y\left(\frac{v_s^2}{\gamma} - 2\frac{v_s\kappa}{|k_n|Q}\right)\tilde{h}_n' \\ &- (\gamma - 1)\left[\frac{\mathrm{i}e_0 A_e}{2}k_y\left(\tilde{h}_{n-1}' + \tilde{h}_{n+1}'\right) + \mathrm{i}k_y\tilde{e}_n'\right]\end{aligned} \tag{4.21}$$

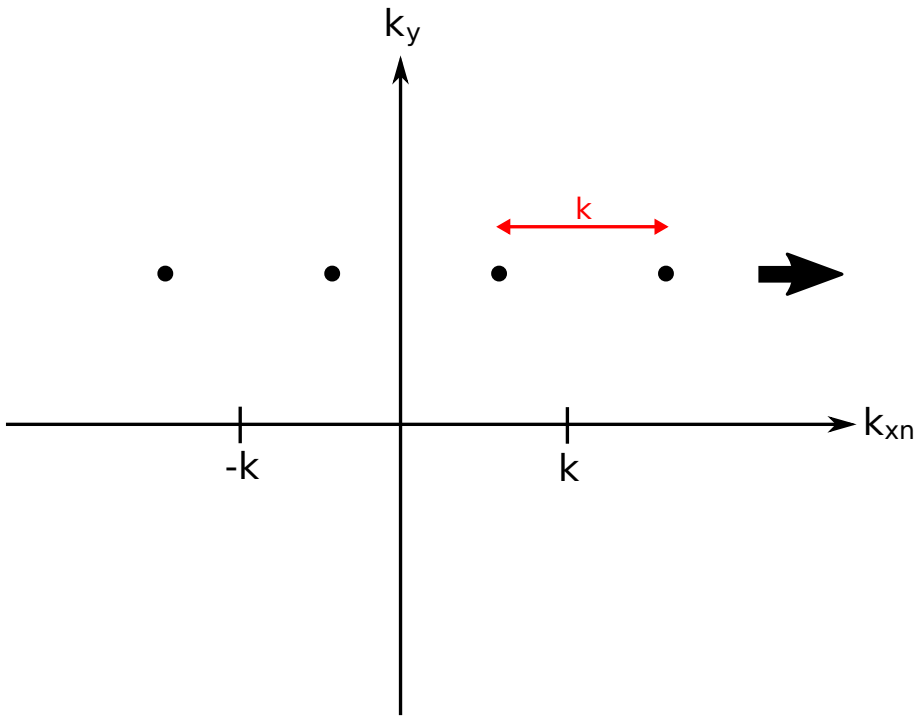


Figure 4.1 – Schematic depiction in Fourier space of the shearing wave ladder prompted by the presence of the zonal flow. The ladder is composed by a number of evenly separated (the separation equal to the axisymmetric structure's wavenumber k) states, which are continuously sheared in the k_{xn} direction dictated by the sign of k_y, as shown by the arrow. The even spacing of the ladder implies that any state will have sheared to exactly the next place in the ladder after some time $t = T_{\rm rec}$. This means that $T_{\rm rec}$ represents the dynamical timescale of the system.

$$\partial_t \tilde{e}'_n = \frac{k e_0 A_e}{2\mathrm{i}} \left(\tilde{u}'_{n-1} - \tilde{u}'_{n+1}\right) - \frac{A_v k_y}{2} \left(\tilde{e}'_{n-1} - \tilde{e}'_{n+1}\right) - \mathrm{i}(\gamma - 1) e_0 \left\{ (k_{xn} \tilde{u}'_n + k_y \tilde{v}'_n) + \frac{A_e}{2} \left[k_{x,n-1} \tilde{u}'_{n-1} + k_{x,n+1} \tilde{u}'_{n+1} + k_y \left(\tilde{v}'_{n-1} + \tilde{v}'_{n+1} \right) \right] \right\} \quad (4.22)$$

where $k_n = \sqrt{k_{xn}^2 + k_y^2}$ and $k_{x,n\pm1} = k_{x0} + (n \pm 1)\, k + q\Omega k_y t$. The system of equations outlined above are valid for both isothermal and adiabatic cases. For the former scenario, all that is needed is to set the value of the adiabatic index to $\gamma = 1$ (compared to $\gamma = 5/3$ for the adiabatic case) and to neglect any contribution caused by e or s, as well as to completely discard Equation 4.22. These simplifications remove one of the system's degrees of freedom, with the imposed zonal flow in

isothermal conditions being described by one parameter only: A_h or A_ζ.

4.2.3 Method

The ladder of shearing waves induced by the presence of the imposed zonal flow upon the background state, as seen in Equation 4.17, comprises an infinite amount of evenly spaced modes, with the separation between any two density wave modes being equal to the wavenumber of the imposed axisymmetric structure as shown in Figure 4.1. It is of course impossible to computationally replicate the infinite ladder, therefore this is truncated to $2N+1$ modes instead (with the chosen range centred around $k_{xn} = 0$). However, the chosen value of N (i.e. the extent of the ladder) could have an impact on the accuracy of the results obtained. An analysis is carried out to better understand how many modes are usually required to describe the shearing wave ladder in a sufficiently accurate way; this is presented in Appendix B. Briefly summarised, the analysis reveals that a ladder of only a few modes is sufficient to provide accurate and converging results. It is of course in the interest of the efficiency of the computational calculation to make use of the smallest possible value of N, without however impacting on the reliability of the results. It is found that values of $2 \leq N \leq 7$ (i.e. the total number of states ranging between $5-15$) provide the best compromise between computational efficiency and result accuracy.

The ladder of shearing waves induced by the presence of the axisymmetric instability possesses a dynamical nature, due to the shear affecting the radial wavenumbers as described by $k_{xn}(t) = k_{x0}(0) + nk + q\Omega k_y t$. This causes the modes in the ladder to shear along the k_{xn} axis, the direction of which is dictated by the sign of k_y (i.e. the modes shear in the negative k_{xn} direction for $k_y < 0$ and viceversa for $k_y > 0$), as indicated by the black arrow in Figure 4.1. Together with the even spacing of the ladder, this implies that there is a well-defined time interval $\Delta t = T_{\text{rec}}$ during which any mode shears by $\Delta k_{xn} = k$, therefore exactly occupying the next place in the ladder at the end of the interval. The existence of the ladder's characteristic timescale

$$T_{\text{rec}} = \frac{k}{q\Omega k_y}, \tag{4.23}$$

which is referred to as the recurrence time of the shearing wave ladder, means that after every integer multiple of the recurrence timescale the ladder would have

modes at the very same k_{xn} values, with only their amplitudes changing due to the ladder being sheared. The calculation presented here, carried out in `MATLAB` and employing Floquet analysis, makes use of this periodicity of the system. $4(2N+1)$ sets of initial conditions (ICs) – in each of which all variables bar one are equal to zero, with the non-zero quantity being different in every set so that each variable is non-zero in one set – are applied to the system described by Equations 4.19–4.22 to form a $4(2N+1) \times 4(2N+1)$ monodromy matrix after said equations are evolved for one recurrence time T_{rec}. Each column of the monodromy matrix is generated by a specific set of the ICs, with the matrix therefore being completed once all $4(2N+1)$ sets of ICs have been applied to the system.

The eigenvalues of the complete monodromy matrix give the factor by which the shearing wave ladder has grown over the evolved recurrence time; the growth rate λ can then be calculated from the eigenvalues using

$$\lambda = \Re\left[\frac{q\Omega k_y}{k}\ln f\right] = \Re\left[\frac{1}{T_{\mathrm{rec}}}\ln f\right], \tag{4.24}$$

where f represents the eigenvalue of the monodromy matrix (i.e. the Floquet multiplier) which gives rise to the maximum growth rate.

It is also worth pointing out that the truncation of the shearing wave ladder (hence the consideration of a finite range of radial wavenumber k_{xn} values), together with the shearing time-dependent nature of k_{xn}, means that modes will eventually shear out of the considered domain; these need to be discarded as they fall out of the domain's bounds, which causes some dissipation in the system. In the same way, new modes shear into the domain and need to be incorporated into the new matrix by shifting the other modes down. Similar procedures are found among non-linear spectral codes (e.g. SNOOPY, see Lesur and Longaretti (2005)).

4.3 Isothermal case

4.3.1 Without self-gravity

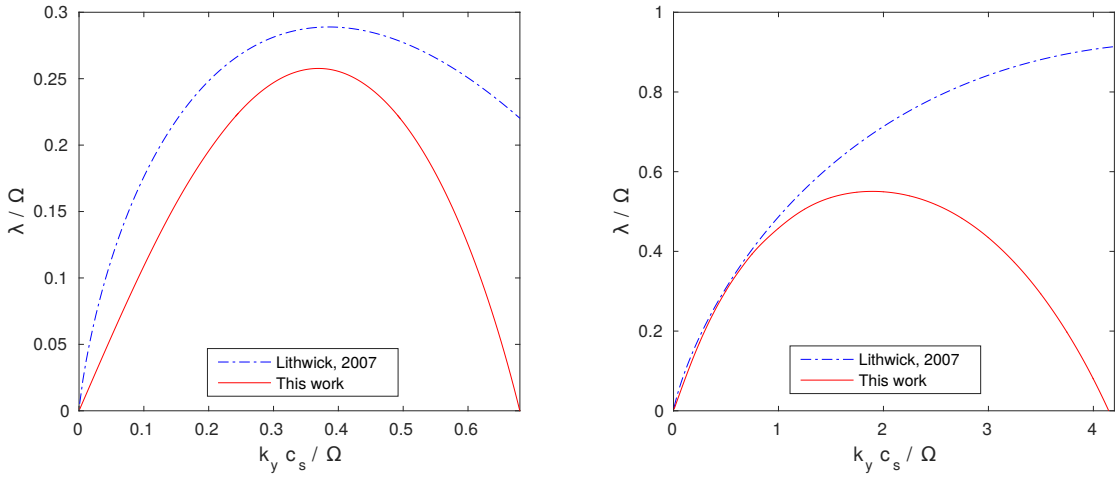
In the non-SG isothermal case the growth rates were found as a function of the azimuthal wavenumber of the disturbance $k_y c_s/\Omega$ and compared to the analysis by Lithwick (2007); as mentioned previously, the work presented in this chapter represents a generalisation of the more simplified problem tackled by Lithwick (2007),

which assumes incompressibility of the flow (the incompressibility condition in the setup described takes the form of $kc_s/\Omega \gg 1$, meaning the imposed axisymmetric structure has a wavelength which is much smaller than the scale height of the disc) and limits the analysis to disturbances that are elongated in the azimuthal direction, so that $k_y/k \ll 1$. With this simplified setup, Lithwick (2007) found the growth rate to be given by

$$\lambda = \frac{q\Omega k_y}{k} \ln\left(\frac{\pi A_h c_s^2 k^3}{4q\Omega^2 k_y}\right), \tag{4.25}$$

where the above expression has been obtained by converting the corresponding equation from Lithwick (2007) into the notation used here.

Comparisons between the growth rates obtained in the analysis presented in this chapter (red, full lines) and that by Lithwick (2007) (blue, dashed lines) are shown in Figures 4.2a and 4.2b for different values of the axisymmetric structure wavenumber.



(a) $kc_s/\Omega = 2, A_h = 0.25$ **(b)** $kc_s/\Omega = 8, A_h = 0.05$

Figure 4.2 – Growth rate λ/Ω as a function of the azimuthal wavenumber obtained in this analysis (red, full line) and by Lithwick (2007) (blue, dot-dashed line) for **(a)**: $kc_s/\Omega = 2$ and $A_h = 0.25$, **(b)**: $kc_s/\Omega = 8$ and $A_h = 0.05$. The results are not in agreement in case **(a)** due to the analysis by Lithwick (2007) being only valid in incompressible conditions ($kc_s/\Omega \gg 1$), which are not met with the current value of $kc_s/\Omega = 2$. In case **(b)** the two curves agree well for $k_y c_s/\Omega \lesssim 1$, where both the incompressibility and $k_y/k \ll 1$ conditions applied in Lithwick (2007) hold.

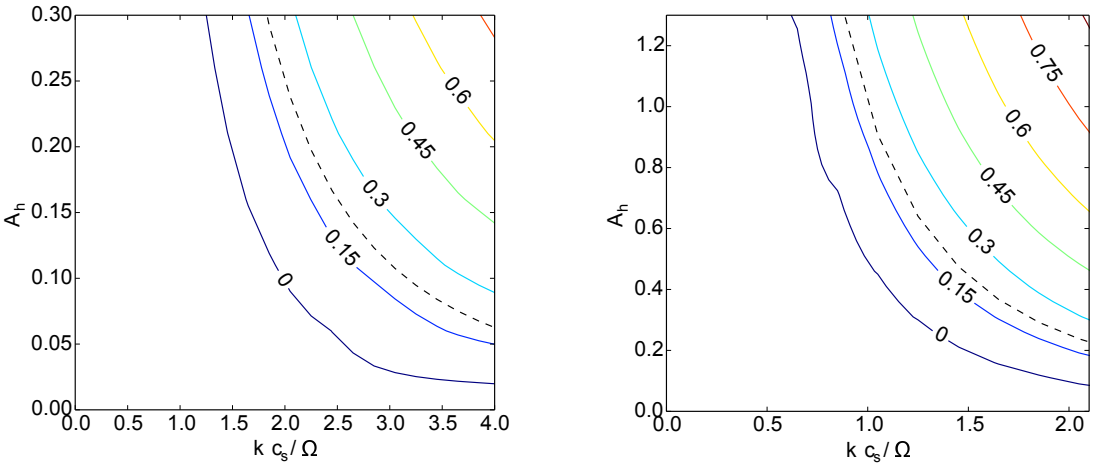
Figure 4.2a, which looks at the case of an axisymmetric structure having wavenumber $kc_s/\Omega = 2$ and amplitude $A_h = 0.25$ shows an inconsistency between the two analyses. This is due to the limitations upon which the work by Lithwick (2007) is based; although the $k_y/k \ll 1$ condition is fulfilled for small $k_y c_s/\Omega$ values, the value of the structure's wavenumber $kc_s/\Omega = 2$ violates the incompressibility assumption ($kc_s/\Omega \gg 1$). In this case Equation 4.25, which requires both of these assumptions, does not hold.

The incompressibility condition is however fulfilled in the case shown in Figure 4.2b, where $kc_s/\Omega = 8$. Here the two curves are in a good agreement for $k_y c_s/\Omega \lesssim 1$ where the $k_y/k \ll 1$ condition also holds. However, as the azimuthal wavenumber is increased this latter assumption is no longer true and the two curves are seen to diverge. Thanks to the excellent agreement between the two analyses in the region where both assumptions hold (i.e. $k_y c_s/\Omega \lesssim 1$), it is possible to identify the instability responsible for the observed growth rates as being of the same type as that found by Lithwick (2007): a Kelvin-Helmholtz (KH) instability.

Stability as a function of the structure's properties

A more thorough stability analysis was carried out to pinpoint how the properties of the imposed axisymmetric structure (i.e. its amplitude and wavelength) affect the stability of the flow. In order to achieve this, the growth rates resulting from a large number of amplitude and wavelength combinations were obtained; of course these growth rates λ would also be dependent on the azimuthal wavenumber k_y of the perturbation, as previously shown in Figures 4.2a and 4.2b. Therefore, the value of λ is maximised with respect to k_y by being computed over roughly 20 evenly separated values of the azimuthal wavenumber, spanning the range $k_y c_s/\Omega \in [0, kc_s/\Omega)$. Choosing the range $k_y c_s/\Omega < kc_s/\Omega$ is justified by Figures 4.2a and 4.2b (as well as other plots not shown in this thesis) which show no growth for $k_y/k \gtrsim 1/2$ and a peak growth roughly around $k_y/k \sim 1/4$. Regardless of the further restrictions to the potential k_y range of interest, it was decided to use the more broader $k_y c_s/\Omega \in [0, kc_s/\Omega)$ domain as a matter of precaution, particularly in the case of low kc_s/Ω values.

The results of this analysis are shown by means of growth rate contours in Figure 4.3a. The strength of the resulting non-axisymmetric instability appears to increase with both increasing zonal flow wavenumber and amplitude; smaller



(a) **(b)**

Figure 4.3 – Growth rate contours as a function of both the axisymmetric structure's amplitude A_h and wavenumber kc_s/Ω, after being maximised over k_y. While the instability induced by the structure becomes stronger with both increasing A and increasing kc_s/Ω, the mere presence of the structure is not a sufficient condition for instability. Indeed Figure 4.3a shows stability up to $A_h = 0.3$ as long as the structure's wavenumber is $kc_s/\Omega \lesssim 1.2$. The area of $kc_s/\Omega \in [0, \sim 2]$ is then further explored in Figure 4.3b for a larger range of A_h to pinpoint whether there exists an absolute threshold value of kc_s/Ω below which the flow is always stable, but it is not possible to tell whether that is the case. In both plots the dashed line represent the marginal stability due to the Rayleigh criterion (i.e. $A_h(kc_s/\Omega)^2 = 2(2-q)\Omega^2$), with the flow being unstable to axisymmetric disturbances on the right of the curve.

wavelength zonal flows are also more likely to trigger an instability, with only a small A_h value required for the system to be unstable if $kc_s/\Omega \gtrsim 3.5$. On the other hand, if the wavelength of the axisymmetric structure is sufficiently long, then even a significantly large amplitude might not be enough to cause an instability. It therefore appears that while a minimum would exist in the potential vorticity profile as a necessary condition to trigger the KH instability (Lin and Papaloizou, 2011a; Lovelace and Hohlfeld, 2013), its mere presence is not a sufficient condition for instability. Instead, a more detailed constraint regarding the shape of said minimum must be considered to establish whether its presence causes an instability; this confirms the results by Lovelace et al. (1999), which found that an instability only occurs if the height-to-width ratio of the potential vorticity minimum exceeds a critical value. The plot also indicates, by means of a dashed line, the set of

points (satisfying the equation $A_h(kc_s/\Omega)^2 = 2(2-q)\Omega^2$) being marginally stable to axisymmetric disturbances according to the Rayleigh criterion; although this type of instability (taking place to the right of the dashed line) cannot operate in the analysis presented as it requires a three-dimensional setup, it is important to keep in mind its presence in a real-life disc. By visualising its operating region by means of the dashed line it is possible to tell whether it would limit the properties of zonal flows forming in an accretion disc. Figure 4.3a clearly shows that – while most of the operating region of the KH instability overlaps with the operating region of the axisymmetric instability described by the Rayleigh criterion – it is possible for the non-axisymmetric instability to operate in a regime that is Rayleigh stable. This means that the zonal flow amplitude and length-scale are ultimately limited by the KH instability.

In the high kc_s/Ω limit, it is possible to deduce the behaviour of the growth rate contours by means of Equation 4.25 from Lithwick (2007); this can be used to infer the value of the zonal flow amplitude A_h for given values of kc_s/Ω and the growth rate λ. This can however only provide an estimate for the behaviour due to the simplifications used in the analysis made by Lithwick (2007) which is apparent in Figure 4.2b, showing a discrepancy between the peak growth rate obtained in this analysis and the corresponding one from Lithwick (2007).

In Figure 4.3b, the analysis is focused on $0 < kc_s/\Omega \lesssim 2$ while considering an extended range of A_h; this is in order to establish whether the threshold value in kc_s/Ω needed to trigger an instability continues to exhibit an A_h dependence or there exists an absolute critical value below which a KH instability is impossible, regardless of the zonal flow amplitude. It is possible to implement the extended range of A_h values, which extends up to $A_h = 1.3$, as the imposed zonal flow is described exactly by the equations provided in the isothermal, non-SG regime. From the data plotted in Figure 4.3b it however appears impossible to establish whether the critical kc_s/Ω value would eventually exhibit asymptotic behaviour or retain a weak A_h dependence.

By constructing a similar plot to Figure 4.3a for a different value of the adiabatic index γ, shown in Figure 4.4, it is possible to show that the threshold kc_s/Ω value is independent of γ. While the change of the adiabatic index value from $\gamma = 1$ to $\gamma = 1.5$ results in a similar plot to that shown in Figure 4.3a, it is possible to spot some minor differences; for example the strength of the instability at any one given point in the plane is somewhat lower in the $\gamma = 1.5$ case. With regards to the

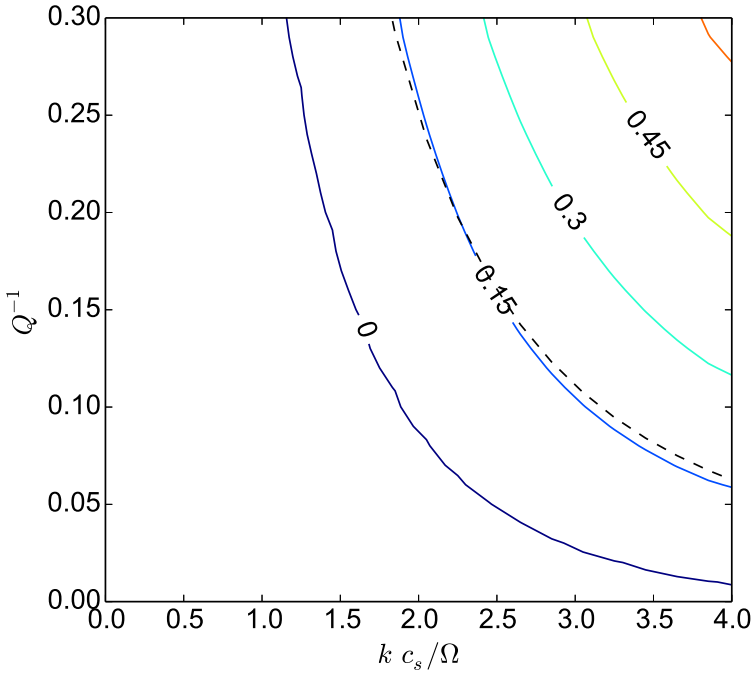


Figure 4.4 – Similar analysis to Figure 4.3a but for $\gamma = 1.5$ rather than $\gamma = 1$. Although the two plots look very similar, it is possible to spot subtle differences between them, such as the analysis for $\gamma = 1.5$ exhibiting a somewhat weaker instability. However the threshold kc_s/Ω value appears to be independent of γ.

kc_s/Ω threshold value, there however appears to be no visible difference in its value between the two cases. The aforementioned threshold can then be thought of as a critical wavelength of the axisymmetric structure, with zonal flows of wavelength $\gtrsim 8H$ being stable to non-axisymmetric disturbances.

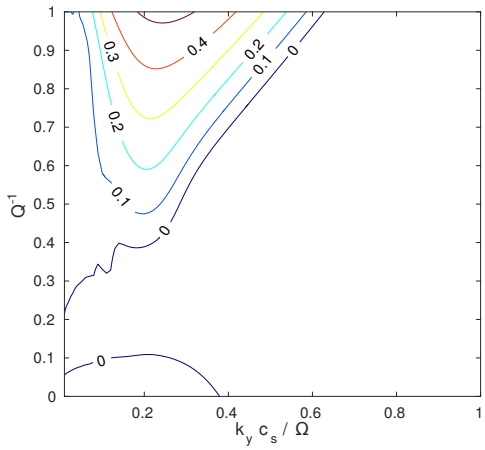
4.3.2 With self-gravity

The analysis, which already represented an extension to the incompressible work by Lithwick (2007), was then further expanded by introducing into the system the effects of self-gravity; this additional component of the disc dynamics could trigger, in suitable conditions, a different type of instability to the one described above. Lovelace and Hohlfeld (2013) carried out a comparable stability analysis in which instabilities were observed both when considering and neglecting self-gravitational

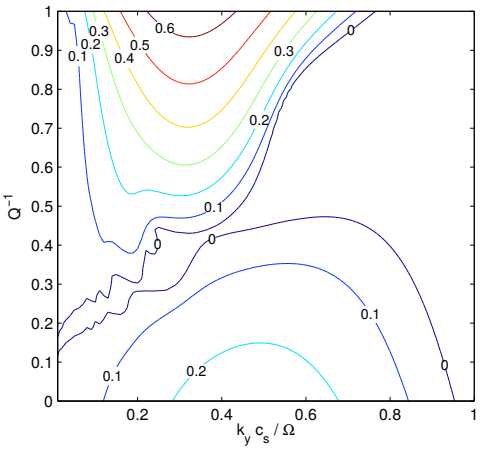
interactions. Their analysis however employed a simpler, single structure featuring a profile of discontinuous nature in the disc density; this is in contrast with the sinusoidal axisymmetric structure in geostrophic balance present in density, velocity and internal energy (if $\gamma \neq 1$, as in Section 4.4) used here.

To investigate whether the introduction of self-gravity introduces a second type of instability, the analysis presented in Figure 4.2 was expanded to calculate the growth rates for a range of self-gravitational strengths (measured by means of the Toomre's parameter Q) as well as for a range of $k_y c_s/\Omega$. This results into a contour plot of the growth rate in the ($k_y c_s/\Omega$, Q^{-1}) plane, with $Q^{-1} = 0$ representing the non-SG case. The plots for $A_h = 0.1$ and wavenumber values ranging from $kc_s/\Omega = 2$ to $kc_s/\Omega = 5$ are presented in Figures 4.5a–4.5d.

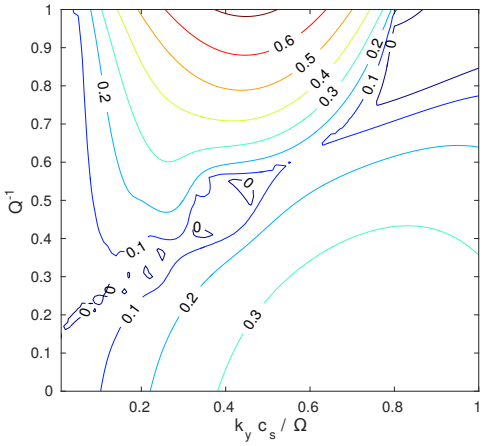
For the $kc_s/\Omega = 2$ case, shown in Figure 4.5a, the previously detected KH instability is again found for $Q^{-1} \lesssim 0.1$; for somewhat higher self-gravity strengths the system is found to be stable, while a non-axisymmetric instability of gravitational nature develops with preferred wavenumber $k_y c_s/\Omega \simeq 0.2$ for $Q^{-1} \gtrsim 0.25$. The $kc_s/\Omega = 3$ case (Figure 4.5b) offers somewhat more detail, with the KH instability seen to prefer disturbances of increasingly shorter wavelengths as its strength wanes when self-gravity is increased away from $Q^{-1} = 0$, with the preferred wavenumber moving from $k_y c_s/\Omega \sim 0.5$ to $k_y c_s/\Omega \sim 0.75$. Likewise the self-gravitational instability is also seen to prefer perturbations of shorter wavelengths as Q^{-1} is increased, eventually settling to $k_y c_s/\Omega \sim 0.35$. An area of marginal stability is again observed in this case, with the gravitational instability stretching only down to $Q^{-1} \sim 0.2$ for very long disturbance wavelengths. The two instabilities are however seen to merge in the $kc_s/\Omega = 4$ (Figure 4.5c) and $kc_s/\Omega = 5$ (Figure 4.5d) cases, with both instabilities preferring higher $k_y c_s/\Omega$ values as kc_s/Ω is increased, particularly the KH instability whose preferred wavenumber in the $kc_s/\Omega = 5$ case is $k_y c_s/\Omega \gtrsim 1$. While the gravitational instability is not seen to stretch much lower than $Q^{-1} = 0.2$ in any of the four cases, the merging of the instabilities is actually due to the KH instability operating at increasingly stronger self-gravity conditions as the wavenumber of the imposed axisymmetric structure is increased, reaching up to $Q^{-1} \sim 0.9$ in the $kc_s/\Omega = 5$ case. The $Q^{-1} > 1$ regime is not explored in this analysis as for such self-gravity strengths the system is axisymmetrically unstable, even without the presence of an imposed zonal flow.



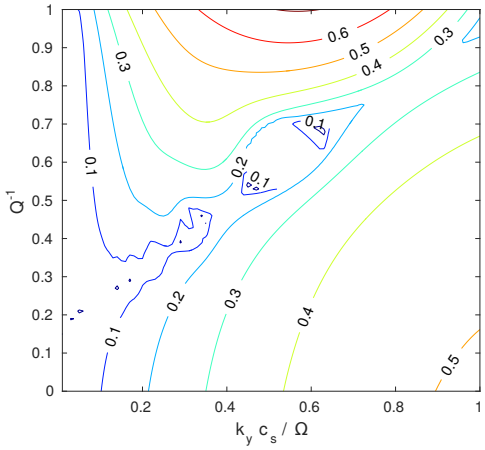
(a) $kc_s/\Omega = 2, A_h = 0.1$



(b) $kc_s/\Omega = 3, A_h = 0.1$



(c) $kc_s/\Omega = 4, A_h = 0.1$



(d) $kc_s/\Omega = 5, A_h = 0.1$

Figure 4.5 – Growth rate contours for $A_h = 0.1$ and $kc_s/\Omega = 2$ (**(a)**, top left), $kc_s/\Omega = 3$ (**(b)**, top right), $kc_s/\Omega = 4$ (**(c)**, bottom left) and $kc_s/\Omega = 5$ (**(d)**, bottom right) in the $k_y c_s/\Omega - Q^{-1}$ plane. For $kc_s/\Omega = 2$ and $= 3$, it appears clear that two distinct types of instability are present in the selected range: a Kelvin-Helmholtz instability for low Q^{-1} and a gravitationally-induced instability for high Q^{-1}. The two instabilities are separated by an area of marginal stability. As the value of kc_s/Ω is raised to 4 and 5, the two instabilities merge. The properties of the instabilities are also seen to evolve when increasing kc_s/Ω, with the gravitational instability stretching to higher $k_y c_s/\Omega$ values, with the most efficient wavenumber similarly moving to shorter wavelengths; the KH instability also stretches to higher $k_y c_s/\Omega$ values, its most effective wavenumber similarly shifting in the same direction from ~ 0.2 ($kc_s/\Omega = 2$) to $\gtrsim 1$.

The nature of the instabilities

After detecting the two instabilities observed in Figure 4.5, the next step was to analyse their behaviour to pinpoint their nature. This is particularly true for the self-gravitating instability, as there have been no means of comparing it with an instability obtained in the literature, unlike the KH instability.

A number of dynamical instabilities taking place in accretion discs are triggered by the presence of an extremum in the potential vorticity profile, or equivalently in the disc surface density (Papaloizou and Pringle, 1985; Papaloizou and Pringle, 1987; Papaloizou and Lin, 1989; Papaloizou and Savonije, 1991; Lovelace et al., 1999; Lin and Papaloizou, 2011a; Lovelace and Hohlfeld, 2013). To determine whether each of the two instabilities is linked to the presence of an extremum in the potential vorticity ζ, and whether this is a maximum or a minimum, the quantity I is introduced. This is the correlation integral between the PV induced by the imposed structure and the wave energy. The latter of these quantities – the wave's acoustic energy – is defined as

$$E = \frac{\Sigma}{2}\left(u'^2 + v'^2\right) + \frac{P'^2}{2\gamma P}, \tag{4.26}$$

which, in the isothermal case, can be simplified into

$$E \propto u'^2 + v'^2 + c_s^2 h'^2. \tag{4.27}$$

The above equation is then azimuthally averaged to give

$$\langle E \rangle_y = \frac{1}{2}\Re\left\{\sum_m \sum_n \left(u'_n u'^*_m + v'_n v'^*_m + c_s^2 h'_n h'^*_m\right) \mathrm{e}^{\mathrm{i}(n-m)kx}\right\}. \tag{4.28}$$

The correlation integral of the above azimuthally-averaged wave energy and the variable part of the potential vorticity ζ results in

$$I \propto -\Re\left\{\sum_n \tilde{u}'_n \tilde{u}'^*_{n+1} + \tilde{v}'_n \tilde{v}'^*_{n+1} + c_s^2 \tilde{h}'_n \tilde{h}'^*_{n+1}\right\}, \tag{4.29}$$

where $\tilde{u}^*$ represents the complex conjugate of $\tilde{u}$ and $\Re$ indicates that only the real part of the expression enclosed within curly brackets is considered. The minus sign has been inserted out of convenience to intuitively pair up positive I values with PV maxima and negative values with minima, rather than viceversa. The time-

average of the correlation integral I for $kc_s/\Omega = 3$ and $A_h = 0.1$ in the ($k_y c_s/\Omega$, Q^{-1}) plane is shown in Figure 4.6.

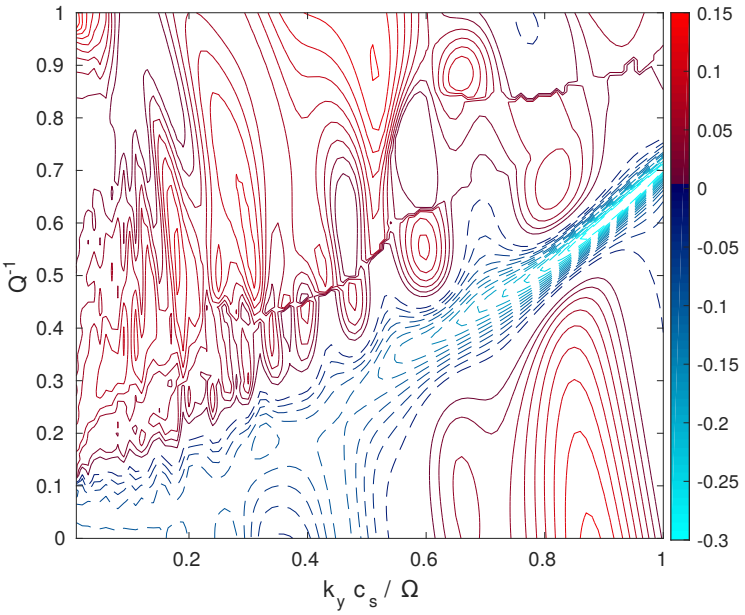


Figure 4.6 – Contour plot of the quantity I, representing the normalised, time-averaged correlation integral of the PV and the wave energy. The plot given is for an axisymmetric structure having $kc_s/\Omega = 3$ and $A_h = 0.1$. Negative (blue, dashed contour lines) and positive (red, full) values of I appear to be mostly divided in two regions of the plane, corroborating the two instabilities observed having different origins. The SG instability appears to be linked to a maximum in the PV ($I > 0$) while the KH instability, apart from a positive region centred around $k_y c_s/\Omega \sim 0.8$, is related to PV minima ($I < 0$).

The plot appears partitioned into two distinct regions, with the top region possessing mostly positive I values (red, full contour lines) and the bottom region being mostly negative (blue, dashed lines). The presence of these two distinctive regions, which are similar to the two instability areas observed in the corresponding growth rate contour plot in Figure 4.5 (Figure 4.5b), confirms the idea of the two instabilities being of different natures. The sign of I indicates that the gravitational instability is centred around a PV maximum, while the KH instability is linked to minima in the potential vorticity. The KH instability region however also presents a rather extended region, covering the range $0.6 \lesssim k_y c_s/\Omega \lesssim 1$, possessing $I > 0$ values, which appears to be linked to PV maxima, instead; this behaviour is explained later.

Taking into account the conclusions drawn from Figures 4.5 and 4.6, it is possible to infer that in non-SG conditions the operating instability (KH instability) is linked to minima in the potential vorticity; introducing self-gravity and making it stronger stabilises the KH instability until the system is in a marginally stable state (for $kc_s/\Omega \lesssim 3$). Strengthening self-gravity further triggers a second type of instability, which this time is centred around a maximum in ζ. This agrees with results of similar stability analyses (Lin and Papaloizou, 2011a; Lin and Papaloizou, 2011b; Lovelace and Hohlfeld, 2013).

Another possibility in pinpointing the nature of the two instabilities is to measure the amount of compressibility and vorticity linked to each instability by means of a contour plot in the ($k_y c_s/\Omega$, Q^{-1}) plane. Given the general characteristics of the two types of instability encountered, one is inclined to expect that the gravitational instability should be linked to a predominance of compressible motions, while the KH instability should exhibit a vortical behaviour. For this purpose, the function below is considered

$$\Upsilon = \ln\left(\frac{\int|\nabla\cdot v|^2\mathrm{d}A}{\int|\nabla\times v|^2\mathrm{d}A}\right), \tag{4.30}$$

which can be converted into Fourier space by means of Parseval's theorem[2] which, for a given real space function $f(x)$, is defined as

$$\int|f(x)|^2\mathrm{d}A = \sum_n|\tilde{f}(\boldsymbol{k})|^2A. \tag{4.31}$$

This transforms Equation 4.30 into the following:

$$\Upsilon = \ln\left(\frac{\sum_n|k_{xn}\tilde{u}'_n + k_y\tilde{v}'_n|^2}{\sum_n|k_{xn}\tilde{v}'_n - k_y\tilde{u}'_n|^2}\right), \tag{4.32}$$

where $n \in [-N, N]$.

The contour plot of Υ obtained for $kc_s/\Omega = 3$ and $A_h = 0.1$, showing both its sign and value, is shown in Figure 4.7. The domain is once again divided into two regions, which roughly agree with those observed in Figure 4.6; as expected, the region corresponding to the gravitationally-induced instability shows a preponderance of compressibility (red, full contours), while the lower KH region features a substantial amount of vortical motions (blue, dashed contours). How-

[2]Parseval's theorem is derived and proved in Appendix C

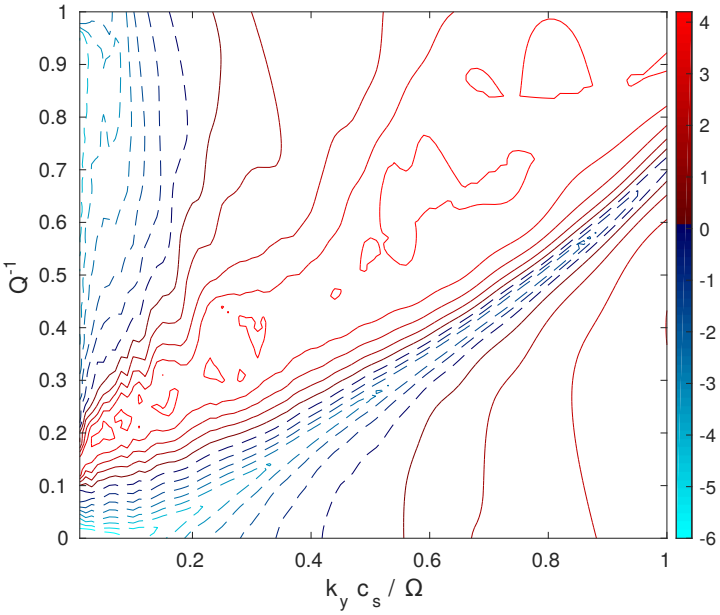


Figure 4.7 – Contour plot of Υ, obtained for $kc_s/\Omega = 3$ and $A_h = 0.1$. The plane is again roughly split in two regions, with the half hosting the KH instability showing a substantial amount of vorticity (blue, dashed contour lines) and the upper half featuring mostly compressive behaviour (red, full lines).

ever, the gravitational instability also presents some vortical behaviour in the low $k_y c_s/\Omega$ regime, highlighting the important role that vorticity (and therefore potential vorticity, too) plays in the problem at hand. Compressive motions on the other hand become more significant (both in the sign and value of Υ) in the limit of $k_y c_s/\Omega \to 1$; in these conditions the production of density waves by means of vortical perturbations is in fact maximised (Heinemann and Papaloizou, 2009a).

Mode visualisation

Unstable modes associated with both KH and gravitational instabilities were then visualised in real space, in order to confirm the findings obtained in Figure 4.6. The real space visualisation was made possible thanks to the use of the Poisson summation formula, given below in its general form:

$$\sum_{n=-\infty}^{\infty} f(x+nL) = \frac{1}{L} \sum_{n=-\infty}^{\infty} \tilde{f}(nk) \exp\left[\mathrm{i}nkx\right]. \tag{4.33}$$

This links the real space function $f(x)$ to its Fourier transform counterpart $\tilde{f}(k)$ by considering an intermediary function $\sum_{n=-\infty}^{\infty} f(x+nL)$; this is a function which includes copies of $f(x)$ being translated by nL, with $L = 2\pi/k$ being the period of $f(x)$. The function $\sum_{n=-\infty}^{\infty} f(x+nL)$ represents the convolution of $f(x)$ with a Dirac comb III_L (i.e. a row of Dirac δ functions) of period L given by

$$\mathrm{III}_L \equiv \sum_{n=-\infty}^{\infty} \delta(x - nL). \tag{4.34}$$

In our particular setup, the Poisson summation formula given above can be applied thanks to the periodicity possessed by the ladder of shearing waves (Equation 4.17), where the period L in this case takes the form of the recurrence time T_{rec} of the ladder, given in Equation 4.23. In this instance, the real space function can be evaluated by means of

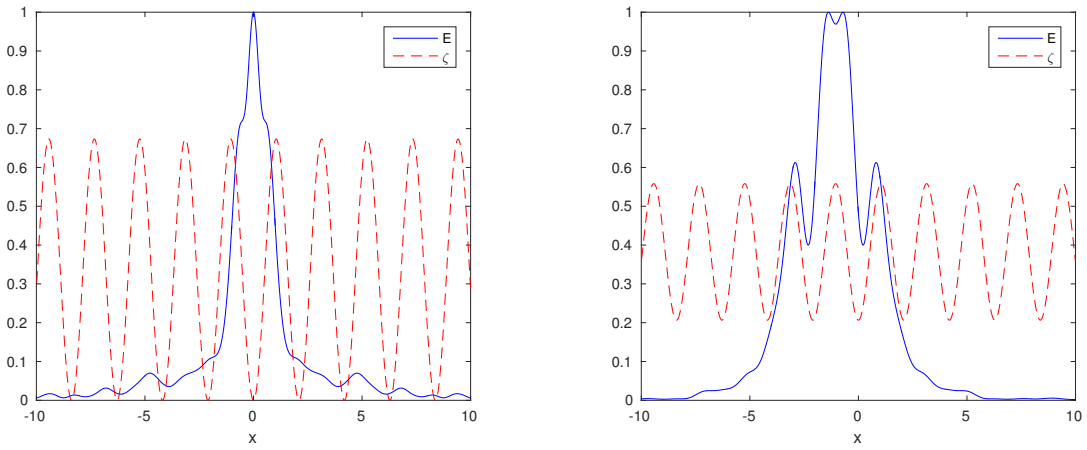
$$\sum_{n=-\infty}^{\infty} f(x+nL) = \sum_{n=-N}^{N} \int_0^{T_{\mathrm{rec}}} \exp(-st)\tilde{f}_n(t) \exp\left[\mathrm{i}\left(n + \frac{t}{T_{\mathrm{rec}}}\right) kx\right] \frac{\mathrm{d}t}{T_{\mathrm{rec}}}, \tag{4.35}$$

where

$$s = \lambda - \mathrm{i}\omega \tag{4.36}$$

is the complex growth rate, with λ being the previously defined real component of the growth rate and ω representing the mode angular frequency.

Equation 4.35 was used to obtain Figure 4.8; this shows the normalised mode energy (blue, full lines) associated with the KH instability (Figure 4.8a) and the gravitational instability (Figure 4.8b) for $kc_s/\Omega = 3$ and $A_h = 0.1$, meaning a direct comparison with Figures 4.5b, 4.6 and 4.7 is possible. The energy plots are accompanied by a suitably scaled real space visualisation of the potential vorticity ζ. This is in order to assess whether the mode of choice is linked to a minimum or maximum in ζ, hereby confirming or rejecting the conclusions reached thanks to Figure 4.6. The mode associated with the KH instability, shown in Figure 4.8a, is obtained for $k_y c_s/\Omega = 0.4$ and $Q^{-1} = 0$ and shows the mode being well localised around a minimum in the potential vorticity, as inferred by Figure 4.6. The mode energy is centred around $x = 0$, which is the location of the mode's co-rotation radius r_{CR}; the value of r_{CR} can be calculated by looking at the Doppler-shifted



(a) **(b)**

Figure 4.8 – Real space normalised energy (blue, full lines) of modes linked with **(a):** KH instability ($k_y c_s/\Omega = 0.4$, $Q^{-1} = 0$) and **(b):** gravitational instability ($k_y c_s/\Omega = 0.3$, $Q^{-1} = 0.9$) obtained for $A_h = 0.1$ and $kc_s/\Omega = 3$. A suitably scaled potential vorticity is also plotted (red, dashed lines), which shows the KH and gravitational instabilities being centred around a PV minimum and maximum, respectively.

angular frequency

$$\begin{aligned}\hat{\omega} &= \mathrm{i}s - k_y(-q\Omega x + v(x)) \\ &= \mathrm{i}s - k_y(-q\Omega x + A_v \sin(kx)).\end{aligned} \tag{4.37}$$

In our setup this can re-expressed, by means of utilising Equation 4.11 in isothermal conditions, into the following:

$$\hat{\omega} = \mathrm{i}s - k_y \left[-q\Omega x + A_h \left(\frac{c_s \kappa}{Q\Omega} - \frac{k c_s^2}{2\Omega} \right) \sin(kx) \right]. \tag{4.38}$$

The location of the co-rotation radius for a neutral mode is where $\hat{\omega} = 0$. For a growing mode, on the other hand, the term $\mathrm{i}s = \omega + \mathrm{i}\lambda$ always has a non-zero imaginary part and the location of co-rotation coincides with the x coordinate satisfying $\Re\{\hat{\omega}\} = 0$. Similarly Figure 4.8b, which is achieved for $k_y c_s/\Omega = 0.3$ and $Q^{-1} = 0.9$, indicates that the gravitationally-induced instability is centred around a maximum in potential vorticity, again as inferred from Figure 4.6. The energy

of the mode is once more centred around the co-rotation radius which, thanks to the equation above, can be pinpointed to be located at $x_{\mathrm{CR}} = -\pi/3$. The energies of both KH and gravitational instability modes also feature density waves which propagate away from x_{CR} and beyond the locations of both Inner and Outer Lindblad Resonances.

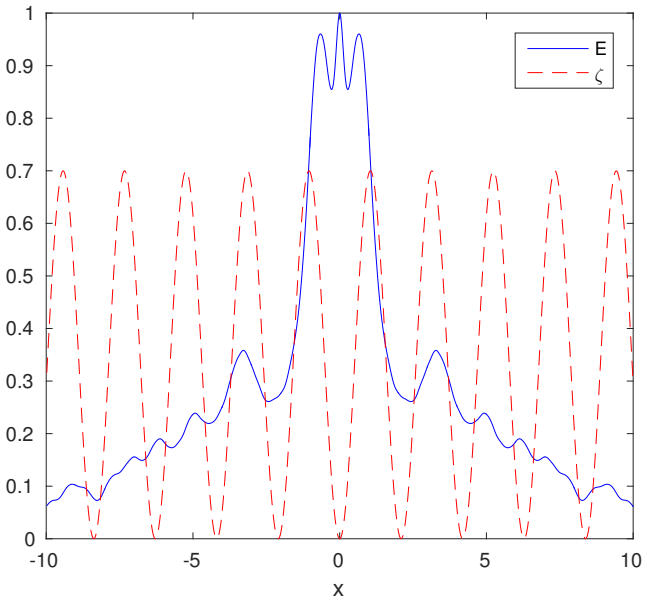


Figure 4.9 – Visualisation of the normalised mode energy in real space associated with the KH instability, similarly to Figure 4.8a. This case however features the mode energy for $k_y c_s/\Omega = 0.7$ and $Q^{-1} = 0$, with this combination of parameters suggesting in Figure 4.6 that the mode should be linked to a PV maximum, rather than minimum. While this mode still appears centred around a PV minimum, its poorer localisation and the stronger secondary peaks due to density wave propagation are thought to give an overall positive value of I.

While the overall trends spotted in Figure 4.6 have been confirmed by the mode visualisation technique employed, there remains the unexpected presence of positive I values within the KH instability region in Figure 4.6, indicating that the modes produced by those combinations of $k_y c_s/\Omega$ and Q^{-1} might be associated with ζ maxima, rather than minima. The situation is looked at in Figure 4.9 which, again illustrating the normalised mode energy and the potential vorticity, is this time obtained for $k_y c_s/\Omega = 0.7$ and $Q^{-1} = 0$. The plot shows that the mode energy is again linked to a minimum in ζ, thereby confirming the result found in Figure 4.8a. However, it appears less localised around the PV extremum

than in the $k_y c_s/\Omega = 0.3$ case, the peak almost enclosing the two neighbouring PV maxima; furthermore, the secondary peaks caused by the outward propagation of density waves also appear enhanced. Both these effects, which are linked to the conclusions drawn from Figure 4.7 – where in the same high–$k_y c_s/\Omega$ region there was an unexpected majority of compressible motions in the KH instability region – are thought to be the cause of the positive I value observed in Figure 4.6.

4.4 Adiabatic case

The more general case of an adiabatic system provides the additional presence of an axisymmetric structure in the entropy, as well as in the potential vorticity (or in e rather than just in h), which could greatly affect the dynamics of the flow and therefore the stability properties analysed in Section 4.3.

The more generalised version of the calculation was first tested by removing the axisymmetric structure in the entropy (i.e. setting $A_s = 0$) in order to confirm the results presented above in Section 4.3. A growth rate contour plot in the ($k_y c_s/\Omega$, Q^{-1}) plane for $\gamma = 5/3$, maintaining $kc_s/\Omega = 3$ and $A_h = 0.1$ as in Figures 4.5b, 4.6 and 4.7, is shown in Figure 4.10. By means of a comparison with Figure 4.5b it is possible to see that the change in the value of the adiabatic index did not induce any major changes. Both KH and gravitational instabilities are present, with the KH instability in particular being very insensitive to the different value of γ. The gravitational instability region, on the other hand, presents small differences compared to its isothermal counterpart; most visibly, the instability appears to stretch to larger $k_y c_s/\Omega$ across all of its operating Q^{-1} range.

The analysis carried out by Lovelace et al. (1999) examined the stability properties of a Keplerian disc in the non-self-gravitating regime; the system was found to be unstable to Rossby wave instability when the radial profile of a specified function – which combines both potential vorticity and entropy – features a maximum. More intriguingly however, a Rossby wave instability was also detected when the aforementioned local maximum was featured in the radial profile of the entropy alone, rather than in the combined function. Lovelace et al. (1999) believe this type of instability can act as a source of outward transport of angular momentum as waves are trapped around the extremum in the entropy.

The adiabatic analysis presented here takes a similar approach to the work by Lovelace et al. (1999) by considering the presence of a zonal flow in both potential

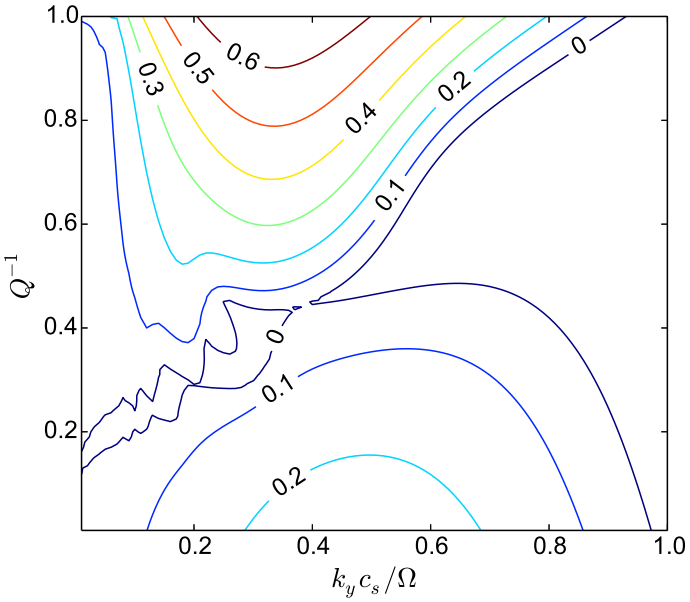
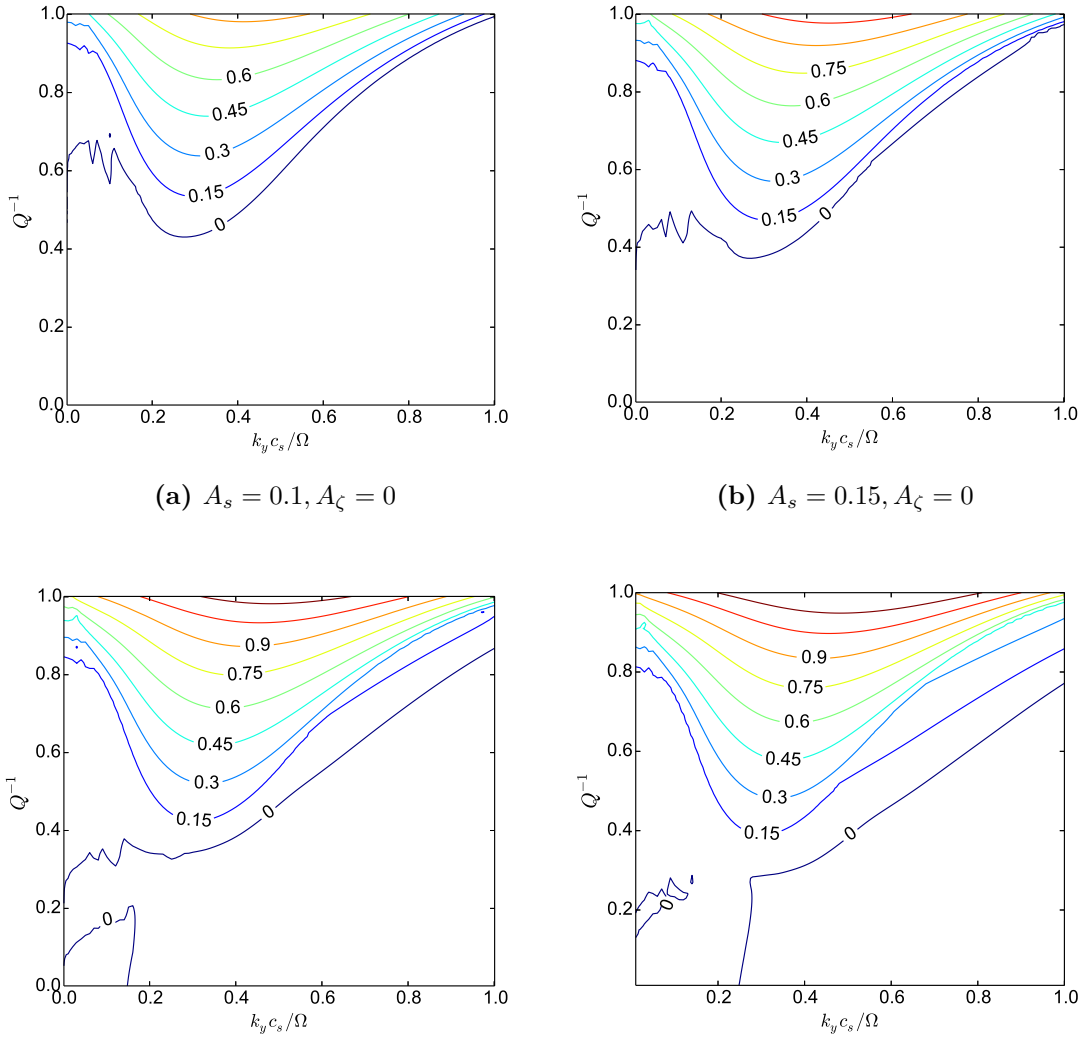


Figure 4.10 – Growth rate contour rate in the $(k_y c_s/\Omega, Q^{-1})$ plane for $\gamma = 5/3$ and $A_s = 0$. A direct comparison with Figure 4.5b is made possible by considering the case with $kc_s/\Omega = 3$ and $A_h = 0.1$. The change in the value of γ from the isothermal case did not cause significant differences. The KH instability region in particular appears unaffected by the new value of γ. The gravitational instability presents some differences, most notably that it extends for a bigger range of wavelengths, but those are of a minor nature.

vorticity and entropy, though it is worth to point out that the entropy in this work is defined somewhat differently to that in Lovelace et al. (1999). The possibility of a Rossby wave instability being triggered when the entropy presents a maximum in its profile is explored below; this is done by setting the amplitude of the potential vorticity structure to zero ($A_\zeta = 0$), while leaving $A_s \neq 0$. The resulting contour plots for the growth rate λ/Ω for a range of A_s values, $\gamma = 5/3$ and $kc_s/\Omega = 2$ are shown in Figure 4.11.

In Figure 4.11a, where the amplitude of the entropy structure is set to $A_s = 0.1$, it is possible to see that setting $A_\zeta = 0$ means that the Kelvin-Helmholtz instability is no longer active. Instead, only a gravitationally-induced instability is triggered, operating for $Q^{-1} \gtrsim 0.4$. As the value of A_s is gradually increased to $A_s = 0.15$, 0.2 and 0.25 in Figures 4.11b, 4.11c and 4.11d, respectively, the gravitational instability stretches down to gravitationally weaker conditions. Finally for $A_s =$



(a) $A_s = 0.1, A_\zeta = 0$ **(b)** $A_s = 0.15, A_\zeta = 0$

(c) $A_s = 0.2, A_\zeta = 0$ **(d)** $A_s = 0.25, A_\zeta = 0$

Figure 4.11 – Growth rate contours in the $(k_y c_s/\Omega,\ Q^{-1})$ plane for $\gamma = 5/3$, $kc_s/\Omega = 2$, $A_\zeta = 0$ and A_s ranging from $A_s = 0.1$ (Figure 4.11a) to $A_s = 0.25$ (Figure 4.11d). As $A_\zeta = 0$ throughout, the KH instability present in the adiabatic regime is no longer active, leaving only a gravitationally-induced instability. As the value of A_s is gradually increased the gravitational instability stretches to progressively lower Q^{-1}, until an instability is detected for $Q^{-1} = 0$ in the $A_s = 0.20$ and 0.25 cases.

0.20 and $A_s = 0.25$ an instability is also observed for $Q^{-1} = 0$, with the two areas of instability merging in Figure 4.11d. It is however unclear whether this represents a unique instability operating in both non-SG and gravitational conditions or two uniformly merged instabilities of different natures.

Although further investigations regarding the nature of this type of instability were beyond the scope of this work, an analysis on how the growth rate yielded by the instability depends on the properties of the imposed axisymmetric structure was conducted. The results, shown in Figure 4.12 for $Q^{-1} = 0$ and $A_\zeta = 0$, shows

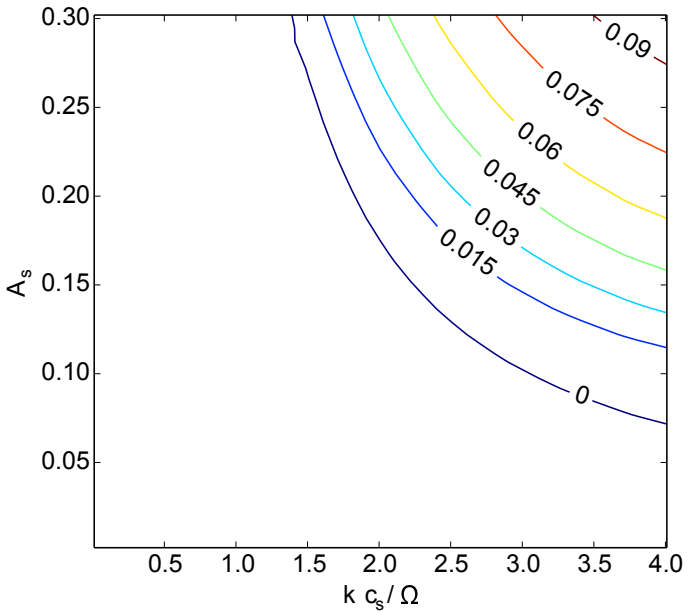


Figure 4.12 – Contour plot for the growth rate λ/Ω maximised over the range $k_y c_s/\Omega \in [0, k)$ for $A_\zeta = 0$ and no self-gravity. Similarly to the KH instability in isothermal conditions, the growth rate strength is seen to grow with increasing kc_s/Ω and increasing A_s, although this instability yields growth rates roughly 1 order of magnitude smaller than those seen in the isothermal regime.

that the instability growth rate increases with both increasing kc_s/Ω and A_s, which is qualitatively similar to the result obtained for the KH instability in Figure 4.3. However the instability seen here yields a growth rate an order of magnitude smaller (of order 10^{-2}) compared to the KH instability in the isothermal regime; moreover, the instability itself operates in a slightly more restricted portion of the plane, with structures with $kc_s/\Omega \lesssim 1.5$ and $A_s \lesssim 0.075$ being stable. Another important difference from the isothermal KH instability is that for $A_\zeta = 0$ the flow is stable

according to the Rayleigh criterion for all values of kc_s/Ω and A_s. This is due to the potential vorticity of the flow being uniform in the absence of a structure in ζ (i.e. $A_\zeta = 0$), causing the vorticity of the system to never change sign.

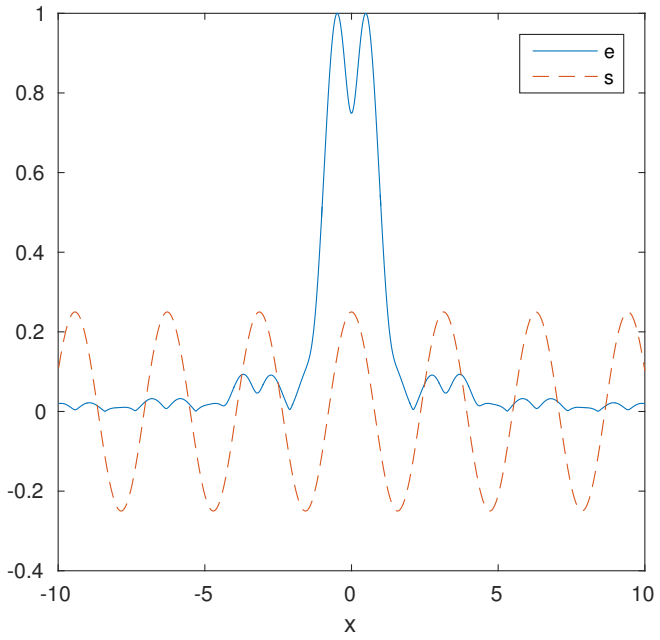


Figure 4.13 – Visualisation of the normalised real space mode energy (red, full line) associated with the instability for $kc_s/\Omega = 2$, $A_\zeta = 0$ and $A_s = 0.25$ in the non-SG regime with $\gamma = 5/3$. A scaled-up version of the entropy s (blue, dashed) is also plotted to highlight the relationship between the two quantities; this shows that the mode is centred around an entropy maximum.

The energy of the instability mode for the entropy-induced instability for $Q^{-1} = 0$ was again visualised using the Poisson summation method outlined in Section 4.3.2. The result, shown in Figure 4.13, is obtained for $\gamma = 5/3$, $kc_s/\Omega = 2$, $A_s = 0.25$, $A_\zeta = 0$ (i.e. the parameter values used in Figure 4.11d), $Q^{-1} = 0$ and $k_y c_s/\Omega = 0.075$. As well as the normalised mode internal energy e (blue, full line), the scaled-up specific entropy s of the system (red, dashed) is also plotted. The energy is well localised around the co-rotation radius of the mode, which occurs at $x = 0$. Moreover, the mode is found to be centred around a maximum in the entropy.

While the magnitude of the growth rates yielded by the entropy-driven instability is smaller (of order of $\lambda/\Omega \sim 10^{-2}$) than the KH instability in isothermal conditions, and might therefore play a secondary role in the stability of the flow,

its very existence is of considerable interest. The current analysis however does not offer further scope of studying the properties of such an instability: the assumption made that the axisymmetric structure's amplitude is small constrains the range of A_s values that can be examined, while higher values of kc_s/Ω than those analysed in Figure 4.12 would imply the wavelengths considered are smaller than H. While it would be of interest to extend the kc_s/Ω range of Figure 4.12 – to establish whether for smaller zonal flow wavelengths an instability is triggered as long as A_s exceeds a critical, wavelength-independent value – this regime would prove inadequate for the 2D model employed.

4.5 Summary

The above stability analysis was based on a steady and ideal Keplerian flow, both neglecting and considering self-gravitational interactions; an axisymmetric sinusoidal structure was also imposed on the flow considered, which caused the flow to be unstable. In the isothermal, non-SG regime the observed growth rates were compared with those obtained by the incompressible calculation by Lithwick (2007); a good agreement was recorded for axisymmetric structures with wavenumber $kc_s/\Omega \gg 1$ and long azimuthal wavelength disturbances, due to limitations in the analysis by Lithwick (2007). The agreement reached between the two works in the relevant conditions means the non-axisymmetric instability observed could be successfully identified as a Kelvin-Helmholtz instability.

The dependence of the instability's growth rate on the properties of the imposed axisymmetric structure (i.e. its amplitude and wavelength) was subsequently studied. The strength of the instability was found to increase with increasing amplitude and decreasing wavelength; however while in the short-wavelength regime an arbitrarily small structure amplitude was necessary to trigger the instability, the system was found to be stable to non-axisymmetric instabilities with structures of amplitudes larger than $\gtrsim 8H$. This threshold wavelength value was found to be independent of the adiabatic index. Zonal flows possessing a wavelength shorter than roughly $8H$ would therefore be unstable to the KH instability, which would break up the existing axisymmetric structure into vortices. As also mentioned in the introduction of this chapter, this instability is however somewhat elusive, so long-lived zonal flows can exist (for instance in MRI simulations) if the box is not sufficiently elongated in the azimuthal direction, as well as in the case of the struc-

ture possessing a stable wavelength. In the former case the work presented here gives an indication that the box aspect ratio required to capture the KH instability must satisfy $L_y/L_x \gtrsim 2 - 2.5$. Although not applicable to the current 2D analysis, the marginal stability conditions dictated by the Rayleigh criterion were also taken into account; the KH instability can however operate in Rayleigh-stable conditions, meaning the properties of the zonal flow are limited by the non-axisymmetric instability. As discussed by Lithwick (2007) and Lithwick (2009), the KH instability causes vortex formation to occur in the disc, although the instability does not appear to directly induce a widening of the axisymmetric structure; this wider, stable zonal flow configuration may however be achieved in an indirect way by means of non-linear dynamics from the formed vortices.

The subsequent introduction of self-gravity triggered a second type of instability, which was found to be distinct and separate from the KH instability for structures with wavelength $\gtrsim 2H$. For smaller wavelength zonal flows, the two instabilities were found to merge as the KH instability operated to increasingly higher values of the inverse of the Toomre parameter Q. By decreasing the structure's wavelength both instabilities were also shifted to shorter operating azimuthal wavelengths. Further analysis by means of a correlation integral unveiled that the KH and gravitational instabilities are centred around minima and maxima in potential vorticity, respectively. The result, which is in agreement with similar works (Lin and Papaloizou, 2011a; Lin and Papaloizou, 2011b; Lovelace and Hohlfeld, 2013) was also confirmed by means of real space visualisations of the energy of modes associated with both instabilities, a method which makes use of the Poisson summation formula. The correlation integral showed however KH modes with $k_y c_s/\Omega \gtrsim 0.5$ seemingly linked with PV maxima instead; this is however a limitation of the technique used as a relevant mode was still found to be centred around PV minimum with the mode visualisation method, although the mode's energy appeared less well localised around its co-rotation radius and secondary peaks were more prominent. An analysis of the amount of vorticity and compressibility associated with each instability was also carried out, with the results generally confirming the initial expectation of the gravitational and KH instabilities possessing a mostly compressible and vortical character, respectively. The gravitational instability however also presented a predominance of vorticity for long $2\pi/k_y$, which highlights the importance of potential vorticity and vortical motions in the problem at hand.

The analysis was then generalised to the adiabatic case by letting the adiabatic

index $\gamma \neq 1$ and by considering internal energy variations. The results obtained in the isothermal case were closely replicated with $\gamma = 5/3$, implying the instabilities detected are insensitive to the value of γ used. Upon removing the structure from the potential vorticity ($A_\zeta = 0$) – leaving one in the entropy – the KH instability was no longer active, while the gravitational instability was still present. Increasing the amplitude of the entropy structure caused said instability to operate at increasingly weaker self-gravitating conditions until, for $A_s = 0.20$, an instability was also detected in non-SG conditions. This agrees with the somewhat surprising results obtained by Lovelace et al. (1999), which detected a Rossby wave instability in the non-SG regime when a local maximum was present in the entropy only. It is however unclear whether the non-SG instability detected is indeed the same as the one operating in the SG regime, or whether there exist two uniformly-merged instabilities. The strength of the non-SG instability was also analysed as a function of the structure's amplitude and wavelength with the results being qualitatively similar to those for the KH instability in the isothermal case; however the adiabatic non-SG instability yields growth rates 1 order of magnitude smaller. The $A_\zeta = 0$ instability also operates in a slightly more restricted range of the structure's properties, with stability obtained for $A_s \lesssim 0.075$ or for structures of wavelength exceeding $\sim 4H$. In this instance the flow is stable according to the Rayleigh criterion for any combination of the structure's property parameters, meaning the physical characteristics of the zonal flow are once more constrained by the non-axisymmetric instability. The Poisson summation method was again used to visualise a mode associated with the instability in real space, with the results showing it to be centred around an entropy maximum.

The analysis presented in this chapter is based on an ideal version of the problem to be tackled: diffusive effects in the disc, as well as cooling effects in the adiabatic case, are in fact neglected; moreover, the imposed axisymmetric structure is assumed to be of a steady nature. These simplifications are of course useful, as they allow the flow dynamics to be analysed directly and clearly, without it needing to be untangled from secondary effects such as those mentioned above. However such an idealised problem represents an unrealistic model of the dynamics of real accretion discs. Further work on this problem would therefore focus on integrating secondary effects to the existing analysis and studying how their introduction affects the stability of the flow.

Although beyond the scope of the current analysis, a further development to the

work presented here would be a deeper study into the nature of the gravitational instabilities encountered in both isothermal and adiabatic regimes. The isothermal instability which counter-intuitively occurs at a density minimum (i.e. PV maximum) is labelled in Lin and Papaloizou (2011a), Lin and Papaloizou (2011b), and Lovelace and Hohlfeld (2013) as 'edge modes', so further work should uncover whether the instability found here is indeed of the same kind. A deeper understanding of the adiabatic gravitational instability is also required, although its growth rate was found to be weaker.

Chapter 5

Zonal flow evolution and overstability in accretion discs

5.1 Introduction

The findings of Chapter 4 indicate that a zonal flow present in the system would be subject to non-axisymmetric instabilities, the nature of which depends on the strength of self-gravity and the zonal flow properties. These instabilities are also seen to operate in Rayleigh stable conditions, meaning they play a central role in constraining the amplitude and wavelength of axisymmetric structures emerging in the flow. While this is of considerable interest due to the role that zonal flows can play in the dynamics of accretion discs, it is important to study how such axisymmetric structures may be formed in the first place, and whether viscous, thermal or self-gravitational effects aid or hinder potential formation mechanisms.

Accretion discs are subject to a wide assortment of instabilities of various types, with the classical thermal and viscous instabilities (e.g. Pringle et al., 1973; Lightman and Eardley, 1974; Shakura and Sunyaev, 1976; Pringle, 1976; Livio and Shaviv, 1977; Piran, 1978; Pringle, 1981) representing two of the most studied examples. The existence of these two types of instability depends on assumptions made regarding the modelling of dissipation and of angular momentum transport, and are therefore distinguished from more fundamental dynamical instabilities such as the MRI, the VSI and the gravitational instability.

A Keplerian disc possessing surface density Σ and angular velocity Ω is in a state of thermal equilibrium if its heating and cooling rates $\mathcal{H}$ and $\mathcal{C}$, which are

given by

$$\mathcal{H} = \frac{9}{4}\nu\Sigma\Omega^2 \propto \alpha T_c \Sigma\Omega, \tag{5.1}$$

$$\mathcal{C} = 2\sigma_{\rm SB} T_{\rm eff}^4 \propto \frac{T_c^4}{\tau_z}, \tag{5.2}$$

are equal. Here $\nu = \alpha c_{\rm iso} H$ represents the kinematic viscosity (where $c_{\rm iso} \propto T_c^{1/2}$ is the isothermal sound speed and $H = c_{\rm iso}/\Omega$ the scale height of the disc), $\sigma_{\rm SB}$ is the Stefan-Boltzmann constant, τ_z (in this instance assumed to be obey $\tau_z \gg 1$) represents the optical depth, and T_c and $T_{\rm eff}$ are the disc central and effective temperatures.

As both viscosity and optical thickness can potentially exhibit a dependence on T_c, the disc is found to be thermally unstable to T_c disturbances if

$$\left.\frac{\partial \ln \mathcal{H}}{\partial \ln T_c}\right|_{\Sigma} > \left.\frac{\partial \ln \mathcal{C}}{\partial \ln T_c}\right|_{\Sigma}; \tag{5.3}$$

this is because, should the cooling and heating rates obey the above expression, an upward (downward) temperature perturbation would induce a runaway heating (cooling). The disc surface density is held constant in the above instability criterion as thermal variations occur on a much shorter timescale than changes in Σ, which are caused by viscous effects; this is illustrated by expressing the thermal timescale $\tau_{\rm th}$ as

$$\tau_{\rm th} \simeq \left(\frac{H}{R}\right)^2 \tau_{\rm visc}, \tag{5.4}$$

where $\tau_{\rm visc}$ is the viscous timescale and $H/R \ll 1$ is the aspect ratio for a thin disc.

While an accretion disc simulated using the popular α model (discussed in Section 2.1.2) appears thermally unstable in its inner regions where the total pressure is dominated by the radiation pressure contribution, it is not clear whether this correlates with observations. In fact, some observations have seemingly proven discs to be thermally stable (e.g. Gierliński and Done., 2004; Done et al., 2007). Competing models also obtain different stability properties for the simulated discs; for instance the implementation of the β model – where the stress is a function of the gas pressure, rather than being a function of the total pressure as in the α model – results in a thermally stable disc (Sakimoto and Coroniti, 1981; Stella and Rosner, 1984; Merloni, 2003). Moreover, although the α model has proven useful

and seems to correlate fairly well with observations it is important to remember its shortcomings, such as neglecting heat transport.

To obtain the criterion for the viscous instability, we consider the steady-state solution $\Sigma(r)$ to the equation describing the evolution of a diffusive disc

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{r}\frac{\partial}{\partial r}\left[r^{1/2}\frac{\partial}{\partial r}\left(\nu\Sigma r^{1/2}\right)\right]. \tag{5.5}$$

The dynamic viscosity $\mu \equiv \nu\Sigma$ is then perturbed such that $\mu = \mu + \delta\mu$. Substituting this into Equation 5.5 gives:

$$\frac{\partial}{\partial t}\left(\delta\mu\right) = \frac{\partial \mu}{\partial \Sigma}\frac{3}{r}\frac{\partial}{\partial r}\left[r^{1/2}\frac{\partial}{\partial r}\left(r^{1/2}\delta\mu\right)\right]. \tag{5.6}$$

The applied perturbation $\delta\mu$ is found to grow if the diffusion coefficient of the above expression is negative. This means a disc is viscously unstable if (Lightman and Eardley, 1974)

$$\frac{\partial\left(\nu\Sigma\right)}{\partial\Sigma} < 0, \tag{5.7}$$

where the derivative is taken at constant r and where the disc is assumed to be both in thermal and hydrostatic equilibrium due to the long timescale on which viscous changes occur.

The classical stability analysis of thermal and viscous instabilities outlined above is however limited to the case of long-wavelength perturbations ($H \ll \lambda_{\rm pert} \ll R_0$, where R_0 is the fiducial radius employed in the shearing sheet model). This causes the two instabilities to be distinct. The analysis presented in this chapter [1] instead considers perturbations of wavelength $\lambda \sim H$, therefore representing an extension to the classical analysis; this also crucially causes the coupling between thermal and viscous instabilities to be non-negligible.

The analysis also deals with the possibility of accretion discs being viscously unstable to axisymmetric oscillations. This type of instability was first outlined by Kato (1978), who found that these oscillation would grow if the viscosity coefficient of the disc turbulence were to increase during compressive motions, hence generating an increased amount of thermal energy. This is a similar mechanism to the one responsible for generating nuclear energy resulting in stellar pulsations.

[1] The content of this chapter is based on Vanon and Ogilvie (2017)

Carrying out a local stability analysis, Kato (1978) established that these oscillations can become overstable if the turbulent viscosity coefficient is a sufficiently fast increasing function of the disc surface density. Since this pioneering work on disc overstability, the regime has also been considered in the context of an α-disc model (Blumenthal et al., 1984) – where it was found that a value of α above a given threshold would result in the oscillations undergoing overstability – as well as in the context of planetary rings and gaseous disc using both linear and non-linear analyses (Kato and Fukue, 1980; Borderies et al., 1985; Papaloizou and Stanley, 1986; Kato et al., 1988; Papaloizou and Lin, 1988; Schmit and Tscharnuter, 1999; Latter and Ogilvie, 2006). This work takes a fresh look at the topic of overstability and applies the idea to self-gravitating discs to analyse how self-gravitating interactions affect this type of instability.

5.2 Thermal balance

The model employed in the analysis mostly coincides with the one presented in Section 3.2. The only difference is a modification to the equation for the temporal evolution of the specific internal energy, whose full form in the current analysis is given below

$$\partial_t e + \boldsymbol{v}\cdot\nabla e = -\frac{P}{\Sigma}\nabla\cdot\boldsymbol{v} + 2\nu_s\boldsymbol{S}^2 + \nu_b(\nabla\cdot\boldsymbol{v})^2 + \frac{1}{\Sigma}\nabla\cdot(\nu_t\Sigma\nabla e) - \frac{1}{\tau_c}(e - e_{\rm irr}). \quad (5.8)$$

The modification affects the last term, where the quantity $e_{\rm irr}$ is introduced; this represents the equilibrium specific internal energy to which the disc would relax in the absence of viscous heat.

In the absence of the $e_{\rm irr}$ term, thermal balance would be given by

$$e_0 = e_0\alpha_s(\gamma - 1)q^2\Omega\tau_c, \quad (5.9)$$

where $\alpha_s = \nu_s\left(\frac{\gamma\Omega}{c_s^2}\right)$ represents the dimensionless shear viscosity parameter. This configuration however only allows for one value of the product $\alpha_s\tau_c$ to satisfy thermal balance, for fixed q and γ. The introduction of $e_{\rm irr}$ instead allows a range of $\alpha_s\tau_c$ values to be allowed, ensuring a more thorough analysis on the stability of the model can be carried out without violating thermal balance. Equation 5.9 is then

altered to

$$e_0 = e_{\rm irr} + e_0 \alpha_s (\gamma - 1) q^2 \Omega \tau_c. \tag{5.10}$$

In the above expression it is possible to identify the fraction of viscously generated heat as

$$f_{\rm visc} = \alpha_s (\gamma - 1) q^2 \Omega \tau_c, \tag{5.11}$$

with its maximum value $f_{\rm visc} = 1$ arising when the disc is heated entirely by viscous effects (i.e. $e_{\rm irr} = 0$).

Assuming that $e_{\rm irr} \geq 0$, Equation 5.10 simplifies to

$$\alpha_s \tau_c \leq \frac{1}{q^2 \Omega (\gamma - 1)}. \tag{5.12}$$

The inequality, which is obeyed in situations of thermal balance, shows that the introduction of $e_{\rm irr}$ allows any value of α_s or τ_c to be considered, as long as their product does not exceed $\frac{1}{q^2\Omega(\gamma-1)}$. The analysis also explores the effect on stability caused by the value of the Prandtl number, which is defined as

$$\Pr = \frac{\nu_s}{\nu_t}. \tag{5.13}$$

Once a perturbation is applied on the background state, the perturbed state being described by $\boldsymbol{v} = \boldsymbol{v}_0 + \boldsymbol{v}'$ (where $\boldsymbol{v}' = (u', v', 0)^T$), its temporal evolution can be described by the set of equations listed below:

$$\partial_t h' = -\partial_x u', \tag{5.14}$$

$$\partial_t u' - 2\Omega v' = -\partial_x \Phi'_{\rm d,m} - (\gamma - 1)\left[\partial_x e' + e_0 \partial_x h'\right] + \left(\nu_b + \frac{4}{3}\nu_s\right)\partial_x^2 u', \tag{5.15}$$

$$\partial_t v' + (2 - q)\Omega u' = \nu_s \partial_x^2 v' - \nu_s q \Omega \partial_x h', \tag{5.16}$$

$$\partial_t e' = -(\gamma - 1) e_0 \partial_x u' - 2\nu_s q \Omega \partial_x v' + \nu_t \partial_x^2 e' - \frac{1}{\tau_c}\left(e' - e_{\rm irr}\right), \tag{5.17}$$

with the analysis being based on the assumption that the cooling τ_c and all three diffusivities ν_i are constant. This can potentially have an effect on the stability of the system described.

As explained in more detail in Section 5.3, the system admits solutions of two kinds: density waves or non-oscillating structures in PV and entropy. These solu-

tions are affected by the viscosities, thermal diffusion and cooling; depending on the values of these quantities and their resulting overall effect, the solutions can be found to be damped, exponentially growing or overstable.

5.3 Evolution

The system described admits standing-wave solutions having an axisymmetric, sinusoidal form such as

$$\begin{aligned} h'(x,t) &= A_h(t)\cos(kx) \\ u'(x,t) &= A_u(t)\sin(kx) \\ v'(x,t) &= A_v(t)\sin(kx) \\ e'(x,t) &= e_0 A_e(t)\cos(kx), \end{aligned} \tag{5.18}$$

where A_h, A_u, A_v and A_e represent the amplitudes of the structure in the respective quantities and k is its wavenumber.

Substituting the above expression into the equations of the system (Equations 5.14–5.17), one can obtain the set of equations describing the temporal evolution of the axisymmetric structure:

$$\partial_t A_h = -kA_u, \tag{5.19}$$

$$\partial_t A_u - 2\Omega A_v = -2\pi G\Sigma_0 A_h + \frac{c_s^2 k(A_e + A_h)}{\gamma} - \left(\gamma_b + \frac{4}{3}\gamma_s\right) A_u, \tag{5.20}$$

$$\partial_t A_v + (2-q)\Omega A_u = -\gamma_s A_v + \gamma_s \frac{q\Omega}{k} A_h, \tag{5.21}$$

$$\partial_t A_e = -(\gamma - 1)kA_u - \gamma_s \frac{2q\Omega}{ke_0} A_v - \gamma_t A_e, \tag{5.22}$$

with $\gamma_b = \nu_b k^2$, $\gamma_s = \nu_s k^2$ and $\gamma_t = \nu_t k^2 + 1/\tau_c$ being three damping coefficients.

Assuming that the above set of equations has exponentially growing solutions of the form $\propto \mathrm{e}^{\lambda t}$, one can obtain a quartic equation for the complex growth rate λ. In the ideal case – where the damping coefficients are ignored – the four solutions to the system are:

$$\lambda_0 = 0, 0, \pm \mathrm{i}\omega_0, \tag{5.23}$$

where the zero subscript denotes the ideal case and the density wave frequency ω_0 is given by

$$\omega_0^2 = \kappa^2 - 2\pi G\Sigma_0|k| + c_s^2 k^2. \tag{5.24}$$

The four solutions correspond to two different types of mode: the zero roots correspond to PV and entropy slow modes, while the non-zero solutions represent density wave modes. The solutions in the ideal case are illustrated in Figure 5.1 by the filled shapes. The introduction of the damping coefficients (which are assumed in the analysis to be small, i.e. $\gamma_i \ll \Omega$) into Equations 5.19–5.22 causes the real part of the modes' growth rates to move away from zero, as shown in Figure 5.1 by means of the empty shapes. If the growth rate's real part is negative the damping coefficients stabilise the mode in question, while if a mode exhibits $\Re(\lambda) > 0$ it will undergo overstability (DW modes) or exponential growth (PV and entropy slow modes).

5.3.1 Regular perturbation

It is however non-trivial to fully understand the effect on the modes' stability caused by the damping coefficients; for non-degenerate solutions (i.e. for the density wave modes possessing $\lambda_0 = \pm \mathrm{i}\omega_0$) it is expected that a regular perturbation analysis should hold. This allows a solution of a problem that cannot be solved exactly to be obtained by means of a power series in a chosen parameter, which is assumed to be small.

In the case of this analysis, the non-ideal growth rate for a given mode with an ideal solution λ_0 subject to a damping coefficient γ_i is given by

$$\lambda = \lambda_0 + \sum_i \gamma_i \lambda_{1,i} + \mathcal{O}(\gamma_i^2), \tag{5.25}$$

where γ_i represents any of the three damping coefficients introduced previously. Indeed the full solution of the problem is bound to depend on each damping coefficient independently, meaning the non-ideal growth rate of a given mode is given by

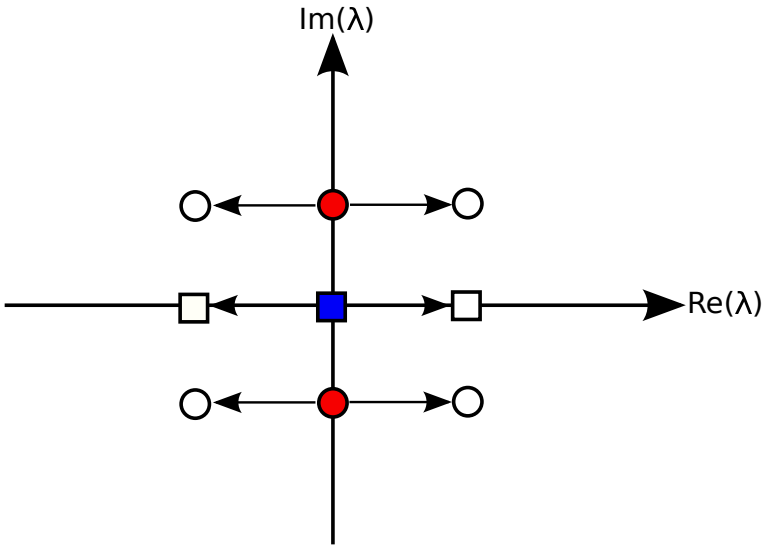


Figure 5.1 – Representation of the solutions' growth rate positions in the Real-Imaginary plane for both ideal (no damping coefficients; filled shapes) and full cases (empty shapes). In the ideal case all solutions possess $\Re(\lambda) = 0$, with the PV and entropy slow modes, also having zero frequency, being indistinguishable from one another (blue square). As the damping coefficients are introduced into the system, the modes' growth rates obtain a non-zero real part; if the newly acquired real part is negative the modes are damped by the non-ideal effects, while if $\Re(\lambda) > 0$ the density wave modes (white circles) are overstable while the slow modes (white square) undergo exponential growth.

$$\lambda \simeq \lambda_0 + \sum_{i=1}^{3} \gamma_i \left(\frac{\partial \lambda}{\partial \gamma_i} \right), \tag{5.26}$$

where the perturbation series was truncated to first order for simplicity.

It was therefore required to obtain an expression for the $\left(\frac{\partial \lambda}{\partial \gamma_i}\right)$ coefficients; this was done by means of the eigenvalue problem

$$\mathbf{M}\boldsymbol{A} = \lambda \boldsymbol{A}, \tag{5.27}$$

where $\mathbf{M}$ represents the matrix containing Equations 5.19–5.22, $\boldsymbol{A}$ represents the eigenvectors of the system, and λ its eigenvalues.

The regular perturbation analysis can, similarly to Equation 5.25, be applied to both $\mathbf{M}$ and $\boldsymbol{A}$ giving

$$\mathbf{M} = \mathbf{M}_0 + \sum_i \gamma_i \mathbf{M}_{1,i}, \tag{5.28}$$

$$\boldsymbol{A} = \boldsymbol{A}_0 + \sum_i \gamma_i \boldsymbol{A}_{1,i} + \mathcal{O}\left(\gamma_i^2\right). \tag{5.29}$$

In the above expressions $\mathbf{M}_{1,i}$ and $\boldsymbol{A}_{1,i}$ represent the additional terms to the system and the eigenvectors resulting from the introduction of a single damping parameter γ_i, while $\mathbf{M}_0$ and $\boldsymbol{A}_0$ represent the matrix and eigenvectors of the ideal case previously discussed, both being known, the former having the form

$$\mathbf{M}_0 = \begin{bmatrix} -k & 0 & 0 & 0 \\ 0 & 2\Omega & -2\pi G\Sigma_0 + c_s^2 k/\gamma & c_s^2 k/\gamma \\ (q-2)\Omega & 0 & 0 & 0 \\ (1-\gamma)k & 0 & 0 & 0 \end{bmatrix}. \tag{5.30}$$

The perturbation series applied to $\mathbf{M}$, $\boldsymbol{A}$ and λ are then applied to the original eigenvalue problem (Equation 5.27), yielding

$$\mathbf{M}_0\boldsymbol{A}_{1,i} + \mathbf{M}_{1,i}\boldsymbol{A}_0 = \lambda_0\boldsymbol{A}_{1,i} + \lambda_{1,i}\boldsymbol{A}_0, \tag{5.31}$$

which holds for each type of diffusivity (i.e. each i) separately, and where terms of second order and above have been neglected and the zeroth-order terms have cancelled by means of

$$\mathbf{M}_0\boldsymbol{A}_0 = \lambda_0\boldsymbol{A}_0. \tag{5.32}$$

Both sides of Equation 5.31 are then multiplied by $\boldsymbol{B}_0^T$, with $\boldsymbol{B}_0$ being the left eigenvector of the ideal case ($\boldsymbol{A}_0$ being the more commonly used right eigenvector), given by

$$\boldsymbol{B}_0^T\mathbf{M}_0 = \lambda_0\boldsymbol{B}_0^T. \tag{5.33}$$

This results in

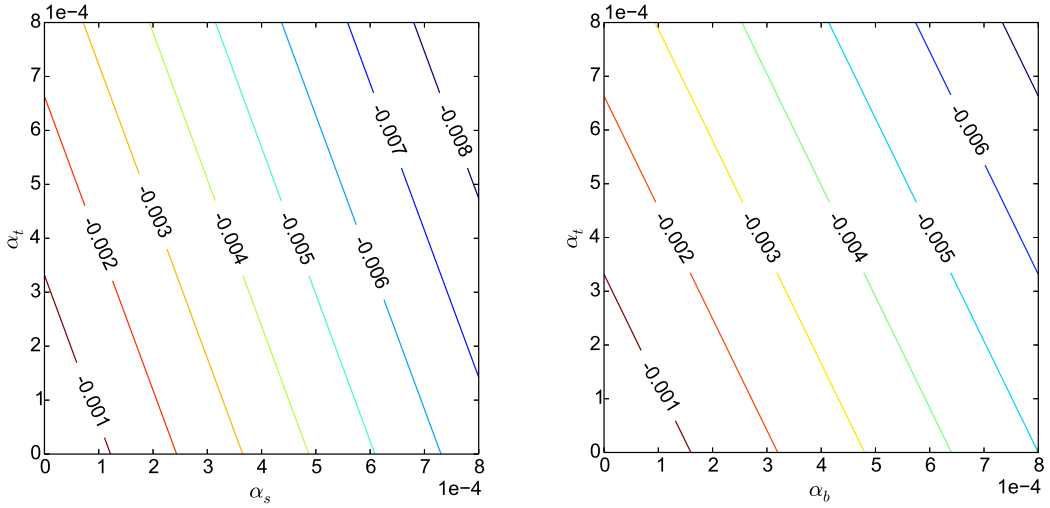
$$\boldsymbol{B}_0^T\mathbf{M}_0\boldsymbol{A}_{1,i} + \boldsymbol{B}_0^T\mathbf{M}_{1,i}\boldsymbol{A}_0 = \boldsymbol{B}_0^T\lambda_0\boldsymbol{A}_{1,i} + \boldsymbol{B}_0^T\lambda_{1,i}\boldsymbol{A}_0, \tag{5.34}$$

which, making use of the relationship expressed in Equation 5.33, finally reduces to the required coefficient $\lambda_{1,i}$ determining the departure from the ideal growth rate λ_0 under the influence of the damping coefficient γ_i:

$$\lambda_{1,i} = \frac{\partial \lambda}{\partial \gamma_i} = \frac{\boldsymbol{B}_0^T \mathbf{M}_{1,i} \boldsymbol{A}_0}{\boldsymbol{B}_0^T \boldsymbol{A}_0}. \tag{5.35}$$

5.3.2 Density waves

The assumption that the growth rates of density waves depend linearly on each type of damping coefficient was found to be appropriate as shown in Figure 5.2, where the growth rates are analysed as a function of all three types of damping coefficients.



(a) **(b)**

Figure 5.2 – Plot showing the DW growth rates as a function of the three diffusivity types, clearly validating the assumption that a regular perturbation may be used in this instance.

The DW growth rates of the full case can then be described by Equation 5.26 and the factors $\frac{\partial \lambda}{\partial \gamma_i}$ are obtained by applying Equation 5.35 for each of the three damping coefficients used; these are found to be

$$\frac{\partial \lambda}{\partial \gamma_b} = -\frac{1}{2}, \tag{5.36}$$

$$\frac{\partial \lambda}{\partial \gamma_t} = -\frac{k^2 c_s^2 (\gamma - 1)}{2\gamma \omega_0^2}, \tag{5.37}$$

$$\frac{\partial\lambda}{\partial\gamma_s} = \left[(\gamma-1)q^2 + 2(2-\gamma)q - 2\right]\frac{\Omega^2}{\omega_0^2} - \frac{2}{3}. \tag{5.38}$$

Analysing the coefficients given above, it is possible to see that the contribution arising from the bulk viscosity is always negative, meaning it is stabilising; γ_t also always produces a stabilising contribution when $\omega_0^2 > 0$ (such that the flow is not dynamically unstable), but the effect due to the shear viscosity is of a more complex nature and establishing its sign is non-trivial. It is also important to remember that, while an individual damping coefficient may provide a destabilising effect on the mode, this would only be unstable if the total contribution $\sum_i \gamma_i \frac{\partial\lambda}{\partial\gamma_i}$ is positive.

An analysis was carried out on the effect on stability caused by the introduction of the shear viscosity damping coefficient γ_s whose contribution represented the most non-trivial instance, as its value and sign are a complicated function of γ, q, Q and k. It can however be pointed out that the expression enclosed within square brackets in the factor $\frac{\partial\lambda}{\partial\gamma_s}$ is positive when considering most value permutations of γ and q which are realistic for the problem at hand. Regions where the shear damping coefficient γ_s would act in a destabilising way are shown in Figure 5.3 for a few values of the adiabatic index γ ranging from $\gamma = 1$ to $\gamma = 2$, under the assumption that $q = 3/2$. The analysis shows that $\gamma = 5/3$ (green, full region) and $\gamma = 2$ (red, dot-dashed) result in overstability contained within $Q \lesssim 2$; this extends for a wide range of kc_s/Ω values, although $kc_s/\Omega \approx 0.5$ appears to be the most unstable one. Decreasing the value of the adiabatic index to $\gamma = 7/5$ (orange, dashed region) and $\gamma = 1$ (blue, dotted) however results in an additional overstable area which stretches to arbitrarily high Q values for sufficiently long wavelengths ($\gtrsim 8H$ and $\gtrsim 16H$, respectively). The overstability regions for $\gamma = 1$ and $\gamma = 7/5$ are consistent with the results obtained by the simplified 2D analysis carried out by Latter and Ogilvie (2006); their work however did not present a stability dependence on the value of γ caused by the shear viscosity, meaning the non-SG overstability is always accessible provided the wavelength in question is adequately long. The cause for the discrepancy between the two analyses is thought to be the absence of a viscous heat modulation in the azimuthal velocity equation in the analysis by Latter and Ogilvie (2006).

Given that – as previously mentioned – the term in square brackets in the expression for $\frac{\partial\lambda}{\partial\gamma_s}$ is usually positive for realistic permutations of q and γ, it is possible to maximise the destabilising effect of overstability due to the shear viscosity by minimising ω_0^2 with respect to k; this can be shown to happen at a wavenumber

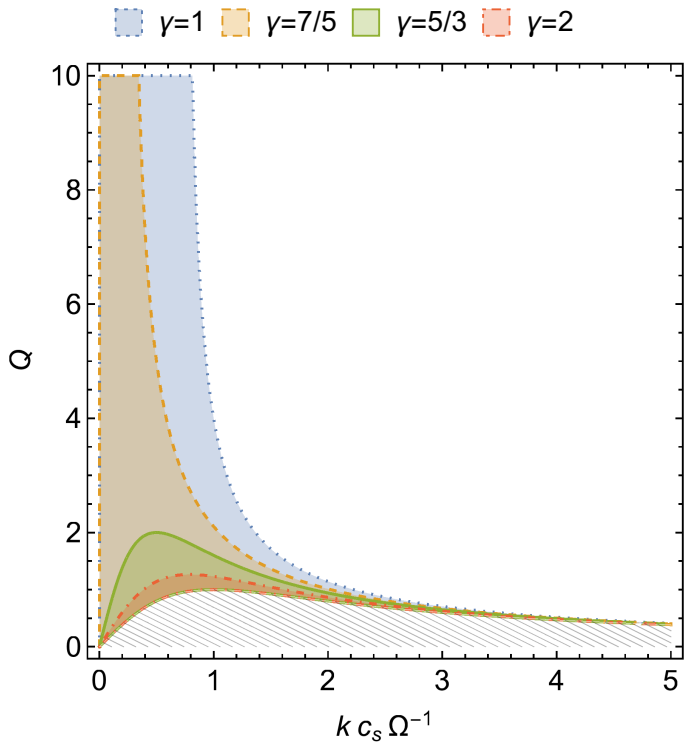


Figure 5.3 – Density wave overstability regions in the kc_s/Ω–Q plane when only the action of shear viscosity is considered, with q assumed to be $q = 3/2$. The shaded regions represent the portion of the parameter space where the shear viscosity contribution has a destabilising effect for a few γ values. While $\gamma = 1$ (blue, dotted region) and $\gamma = 7/5$ (orange, dashed) cause a non-SG instability, higher values such as $\gamma = 5/3$ (green, full) and $\gamma = 2$ (red, dot-dashed) only present a destabilising region for $Q \lesssim 2$. The area corresponding to $\omega_0^2 < 0$, represented by grey hatching, is dynamically unstable and has been neglected.

$$k_{\max} = \frac{\pi G \Sigma_0}{c_s^2}. \tag{5.39}$$

This implies that the minimised value of ω_0^2 is given by

$$\left(\omega_0^2\right)_{\min} = \kappa^2 \left(1 - \frac{1}{Q^2}\right) = 2(2-q)\Omega^2 \left(1 - \frac{1}{Q^2}\right). \tag{5.40}$$

By substituting $(\omega_0^2)_{\min}$ back into the expression for $\frac{\partial \lambda}{\partial \gamma_s}$ one obtains the critical value of the adiabatic index γ, as a function of q and Q, needed for the shear viscosity contribution to be destabilising. The system is found to be susceptible to overstability if

$$\gamma < \frac{2-(2-q)^2-\frac{4}{3}(2-q)\left(1-\frac{1}{Q^2}\right)}{q(2-q)}, \tag{5.41}$$

provided $Q \geq 1$. In the weak- or non-SG limit the coefficient $(1-1/Q^2) \to 1$, which reduces Equation 5.41 to

$$\gamma < \frac{16q-3q^2-14}{3q(2-q)}. \tag{5.42}$$

Assuming $q = 3/2$ for a Keplerian disc, shear viscosity is found to play a destabilising role to overstability for $\gamma < \gamma_{\text{crit}} \approx 1.444$.

The next step was to introduce the damping coefficient γ_t – dependent on both thermal diffusion and cooling – such that the effects on stability due to both γ_s and γ_t are considered; the contribution due to bulk viscosity (i.e. γ_b) was however neglected due to its lack of dependence on the system's parameters. In this instance Equation 5.26 was applied to establish whether the overall effect was stabilising or destabilising with respect to density wave modes. Analytically it can be shown that overstability under the influence of γ_s and γ_t would develop if

$$\frac{\gamma_t}{\gamma_s} < \frac{2\gamma\Omega^2\left[(\gamma-1)q^2+2(2-\gamma)q-2-\frac{2}{3}\omega_0^2/\Omega^2\right]}{k^2c_s^2(\gamma-1)}, \tag{5.43}$$

which can be simplified to

$$\frac{\gamma_t}{\gamma_s} < -\frac{5\left[4k\left(kc_s^2-2c_s\kappa/Q\right)+\Omega^2\right]}{6k^2c_s^2}, \tag{5.44}$$

by letting $q = 3/2$ and $\gamma = 5/3$, where the ratio ω_0^2/Ω^2 has been expanded in the latter inequality to express it as a function of only k and Q. The two expressions highlight the stabilising effect of the damping coefficient γ_t (i.e. of thermal diffusion and cooling), with overstability being triggered only if the damping coefficient ratio γ_t/γ_s is smaller than a critical value, which is a function of q, Q, k and γ. The cooling in particular, as it is the only component of the damping coefficients $\gamma_t = \nu_t k^2 + 1/\tau_c$ and $\gamma_s = \nu_s k^2$ to be independent of k, plays a dominant (stabilising) role in the long-wavelength limit, which, as seen in Figure 5.3, is the preferred limit for overstability, particularly in the non-SG regime. Considering the $k \to 0$ limit in the non-SG regime, as well as the coupling between shear viscosity and cooling prompted by thermal balance (Equation 5.12), one finds the real part of

the overstability growth rate to be given by:

$$\Re(\lambda) = \gamma_s \frac{\Omega^2}{6\kappa^2}\left[-28 + 4(8-3\gamma)q + 6(\gamma-1)q^2 - 3(\gamma-1)^2 q^2 \frac{1}{f_{\rm visc}}\right]. \qquad (5.45)$$

From this expression, a choice of q and of the fraction of heat that is viscously generated, $f_{\rm visc}$, leads to the maximum value of γ for which such a non-SG overstability is possible. By considering a Keplerian disc ($q = 3/2$) which is fully heated by viscous effects ($f_{\rm visc} = 1$) it can be found that non-SG overstability would ensue if the adiabatic index obeys $\gamma \lesssim 1.305$. This represents a stricter constraint than that obtained in the case where only the γ_s damping coefficient was considered, once again highlighting the stabilising effect of γ_t on overstability.

The validity of the threshold value of γ for non-SG overstability is tested in Figure 5.4, where the overstable regions in the kc_s/Ω–Q plane are obtained for $\gamma = 1.3$ and $\gamma = 1.4$, for a Prandtl number $\mathrm{Pr} = \alpha_s/\alpha_t = 5$ (with a shear viscosity of $\alpha_s = 0.05$), and $q = 3/2$. The stability analysis is carried out for a range of cooling times permitted by thermal balance (Equation 5.12), where the longest cooling time used in each case corresponds to a disc being heated almost entirely by viscous effects. In the $\gamma = 1.3$ case, overstability is triggered for all cooling times considered, with the overstable region shrinking as τ_c is shortened, highlighting the stabilising effect of cooling; weak-SG/non-SG overstability is observed for wavelengths longer than $\sim 16H$ for the two longest cooling times: $\tau_c = 20\Omega^{-1}$ (green, full lines) and $\tau_c = 29\Omega^{-1}$ (red, dot-dashed). As γ is increased to $\gamma = 1.4$, which is above the maximum γ estimate for non-SG found above ($\gamma_{\rm crit} \approx 1.305$), non-SG overstability is no longer observed. Furthermore, the shortest cooling time ($\tau_c = 5\Omega^{-1}$) stabilises the system completely and no overstability is observed. The hatched region, which corresponds to $\omega_0^2 < 0$ and where the system is therefore dynamically unstable to axisymmetric perturbations, is neglected.

The maximum value of Q over all kc_s/Ω as a function of q and γ, under the influence of γ_s and γ_t, can be derived analytically by manipulating Equation 5.43 to obtain an expression of the form

$$\frac{1}{Q} = ak + \frac{b}{k}. \qquad (5.46)$$

This can then be maximised over k, with the expression for the wavenumber cor-

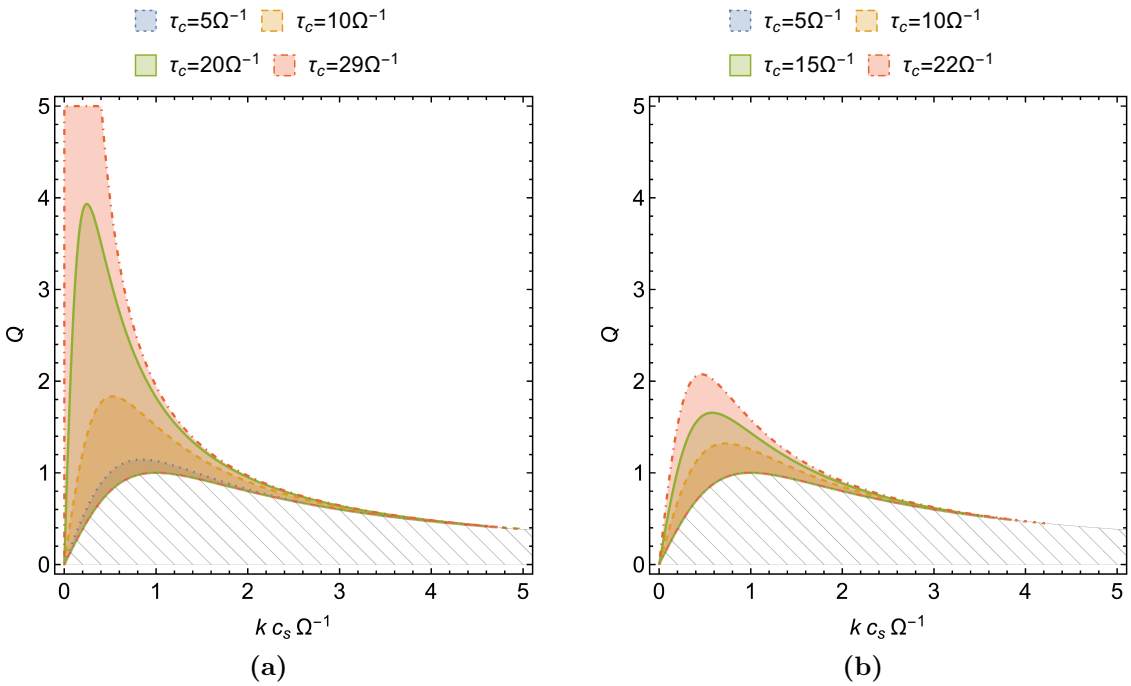


Figure 5.4 – Overstability regions in the kc_s/Ω–Q plane under the influence of shear (γ_s) and thermal (γ_t) damping coefficients for **(a)** $\gamma = 1.3$ and **(b)** $\gamma = 1.4$. The Prandtl number is $\mathrm{Pr} = 5$ (with $\alpha_s = 0.05$) and $q = 3/2$. The stability analysis is carried out for a range of cooling times permitted by thermal balance. For $\gamma = 1.3$ overstability is triggered for all cooling times considered, and a weak-SG/non-SG overstability is observed for the two longest cooling times: $\tau_c = 20\Omega^{-1}$ (green, full lines) and $\tau_c = 29\Omega^{-1}$ (red, dot-dashed). As γ is increased to 1.4 (so beyond the threshold value of $\gamma \approx 1.305$) weak-SG overstability no longer occurs, it being instead constrained to $Q \lesssim 2$; moreover the shortest cooling time for $\gamma = 1.4$ ($\tau_c = 5\Omega^{-1}$) does not trigger any overstability. The hatched region at small Q represents the region where the density wave frequency $\omega_0^2 < 0$ and the system is therefore gravitationally unstable to axisymmetric instabilities.

responding to the maximum Q substituted back into Equation 5.46 to give

$$\frac{1}{Q^2_{\mathrm{max}}} = \frac{\left[4\gamma + 3(\gamma-1)\mathrm{Pr}^{-1}\right]}{32(2-q)\gamma}\left\{28 - 4(8-3\gamma)q - 6(\gamma-1)q^2 + 3(\gamma-1)^2q^2\frac{1}{f_{\mathrm{visc}}}\right\}. \tag{5.47}$$

Assuming that the sum of the first three terms in the curly braces is greater than zero (as if it were negative the system might be overstable in the non-SG

regime and there would not exist a critical Q value), $Q_{\max}$ can be seen to be an increasing function of both $f_{\rm visc}$ and Pr. An example of Equation 5.47 is represented

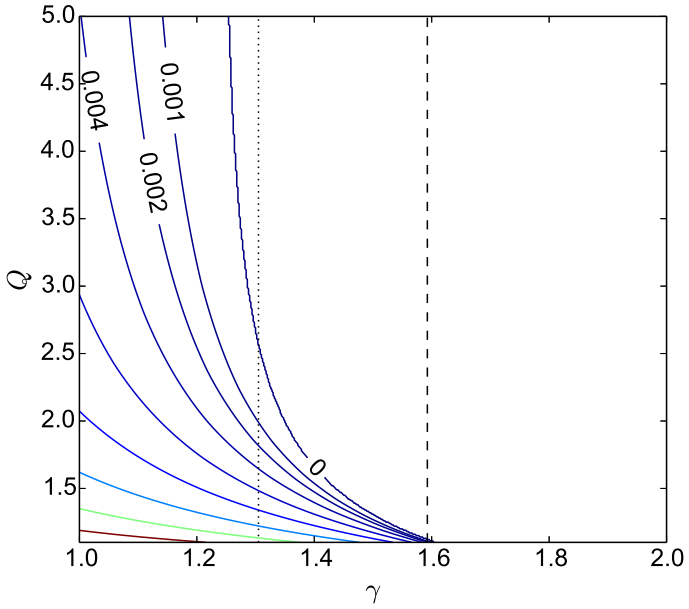


Figure 5.5 – Growth rates maximised over k showing the values of the Toomre parameter Q subject to overstability as a function of the adiabatic index γ, for $q = 3/2$ and $\tau_c = 15\Omega^{-1}$. The values of the Prandtl number $\mathrm{Pr} = 5$ and of the dimensionless shear viscosity $\alpha_s = 0.05$ match those used in Figure 5.4. While a wide range of γ values can trigger overstability for $Q \sim 1$, only values below $\gamma \approx 1.25$ can provide overstability if $Q \sim 5$. The dotted vertical line illustrates the estimated critical value $\gamma \approx 1.305$ above which a non-SG overstability cannot be triggered, while the dashed line represents the largest γ value ($\gamma \approx 1.6$) satisfying the thermal balance condition, with greater values not permitted.

in Figure 5.5, where the range of overstable Q values is explored as a function of γ by means of growth rates optimised over k. The plot, obtained for $\mathrm{Pr} = 5$, $\alpha_s = 0.05$, $\tau_c = 15\Omega^{-1}$ and $q = 3/2$ shows that adiabatic index values up to $\gamma \approx 1.6$ are overstable in strong-SG conditions ($Q \sim 1$), but the largest γ needed for overstability gradually decreases with increasing Q, with only $\gamma \lesssim 1.25$ triggering an instability in weak-SG conditions ($Q \sim 5$). This behaviour is in agreement with the estimated maximum γ value for non-SG overstability ($\gamma \approx 1.305$), which is indicated by means of a dotted vertical line. The plot also features a dashed vertical line, which instead represents the largest γ value permitted by thermal balance for the chosen cooling time.

The introduction of the bulk damping coefficient γ_b – which has a constant, stabilising effect on density waves – further complicates the overstability criterion, its full form (for a choice of $q = 3/2$ and $\gamma = 5/3$) being

$$\gamma_b < -\frac{\left[6k^2c_s^2\right]\gamma_t + \left[5\left(4\omega_0^2 - 3\right)\Omega^2\right]\gamma_s}{15\omega_0^2}. \tag{5.48}$$

5.3.3 Slow modes

The regular perturbation method outlined in Section 5.3.1 cannot however be employed for the slow potential vorticity and entropy modes, whose solutions in the ideal case are coinciding and degenerate. The real parts of their growth rates are in fact found to depend non-linearly with the damping coefficients γ_s and γ_t as shown in Figure 5.6 for $q = 3/2$, $\gamma = 5/3$, $kc_s/\Omega = 2$ and $Q = 1.2$. The modes are also seen to interfere – such that their growth rates form a complex conjugate pair – for a shear damping coefficient $\gamma_s \lesssim 0.03$. Instability is also observed for one of the two modes in a substantial fraction of its plot; instability is strongest in the high–γ_s, low–γ_t regime, showing the thermal damping coefficient γ_t also has a stabilising effect on slow modes. Combinations of γ_s and γ_t falling below the dashed lines in the plots are not allowed as they do not obey thermal balance.

As the regular perturbation method cannot be used for these modes, a set of equations for the evolution of the structure's amplitude in potential vorticity and entropy (A_ζ and A_s) of the form

$$\partial_t A_s = c_1 A_s + c_2 A_\zeta, \tag{5.49}$$

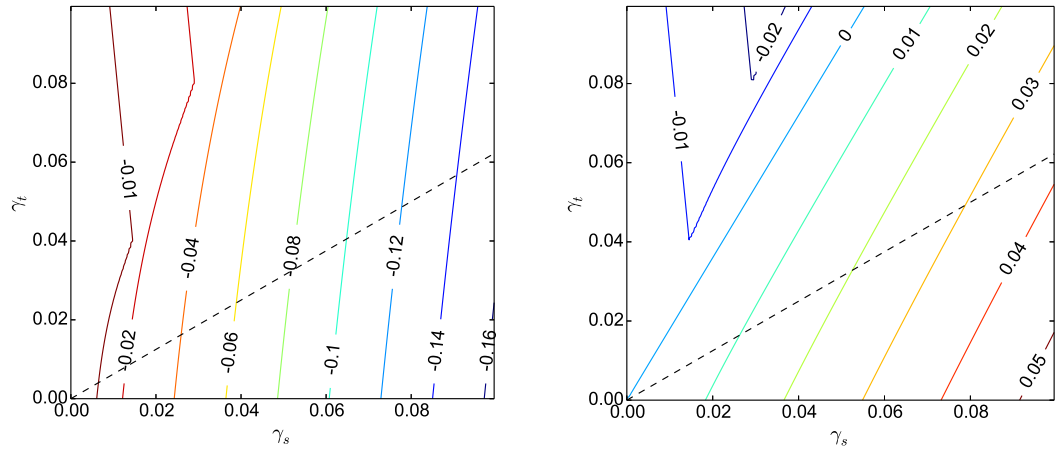
$$\partial_t A_\zeta = c_3 A_s + c_4 A_\zeta, \tag{5.50}$$

are derived analytically, where the coefficients c_1, c_2, c_3 and c_4 are independent of A_u, A_v, A_h and A_e (hence only depend on A_s and A_ζ, as well as the general flow parameters) and where A_s and A_ζ are given by

$$A_s = \frac{1}{\gamma}(A_e + A_h) - A_h, \tag{5.51}$$

$$A_\zeta = \frac{kA_v}{(2-q)\Omega} - A_h. \tag{5.52}$$

The full form of the coefficients is found to be given by:



(a) **(b)**

Figure 5.6 – Slow mode growth rates as a function of γ_s and γ_t for $Q = 1.2$, $kc_s/\Omega = 2$ and $\gamma = 5/3$; both modes present non-linearities in their behaviour, with an interference between them seen for $\gamma_s \lesssim 0.03$, therefore confirming the regular perturbation method cannot be applied. The dashed lines represent the thermal balance limit, with combinations of γ_t and γ_s falling below the lines not allowed.

$$c_1 = \frac{\gamma_t \left(c_s^2 k^2(\gamma - 1) - \gamma\omega_0^2\right) + \gamma_s q \kappa^2 \gamma(\gamma - 1)}{\gamma\omega_0^2}, \tag{5.53}$$

$$c_2 = \frac{\kappa^2(\gamma - 1)\left[\frac{\gamma_t c_s^2}{\gamma} + \frac{\gamma_s q}{k^2}\left(\kappa^2 - \omega_0^2\right)\right]}{c_s^2\omega_0^2}, \tag{5.54}$$

$$c_3 = -\frac{4\gamma_s c_s^2 k^2(q - 1)\Omega^2}{\kappa^2\omega_0^2}, \tag{5.55}$$

$$c_4 = -\frac{\gamma_s\left(\omega_0^2 + 4(q - 1)\Omega^2\right)}{\omega_0^2}. \tag{5.56}$$

Assuming the modes have an exponential solution, the eigenvalue method can be used to obtain the following generic quadratic expression for the growth rate λ of Equations 5.49–5.50:

$$\lambda^2 - (c_1 + c_4)\lambda + c_1 c_4 - c_2 c_3 = 0, \tag{5.57}$$

whose solution can readily be shown to be

$$\lambda = \frac{(c_1 + c_4)}{2} \pm \sqrt{\frac{(c_1 - c_4)^2}{4} + c_2 c_3}. \tag{5.58}$$

In order to check whether a certain set of parameters give rise to an instability, it is possible to either check whether $\Re(\lambda) > 0$ or apply an appropriate stability condition; both of these routes were explored in the analysis, also as a means of confirming the results. For the latter option, the Routh-Hurwitz stability criterion was applied; this represents both the necessary and sufficient stability condition for a linear time-invariant system possessing a polynomial characteristic equation. This states that a second order polynomial having the generic form

$$x^2 + a_1 x + a_0 = 0, \tag{5.59}$$

is stable if each and every coefficient is positive (i.e. $a_i > 0$). For our generic quadratic equation in the growth rate λ, this is equivalent to

$$a_1 = -(c_1 + c_4) > 0, \tag{5.60}$$

$$a_0 = c_1 c_4 - c_2 c_3 > 0, \tag{5.61}$$

with the system being stable only if both conditions are obeyed.

In both long-wavelength ($kc_s/\Omega \to 0$) and short-wavelength ($kc_s/\Omega \to \infty$) limits (the former regime found stable according to the classical theory), the product of the coupling coefficients $c_2 c_3$ results negligible compared to $c_1 c_4$, which represents the product of the respective growth rates of entropy and PV. Furthermore, as both c_1 and c_4 are negative in the long- and short-wavelength limits, both Routh-Hurwitz stability criteria are obeyed and the system is therefore stable. The intermediate-wavelength regime offers however a more complicated picture, as here the coupling coefficients are no longer negligible compared to c_1 and c_4; this means that entropy and PV are coupled, leading to more complex stability properties which are beyond a purely analytical approach. Furthermore, the importance of the coupling coefficients results in the system being potentially unstable even if both c_1 and c_4 coefficients are negative (like in the long- and short-wavelength regimes), as the $c_1 c_4 - c_2 c_3 > 0$ stability condition could be violated.

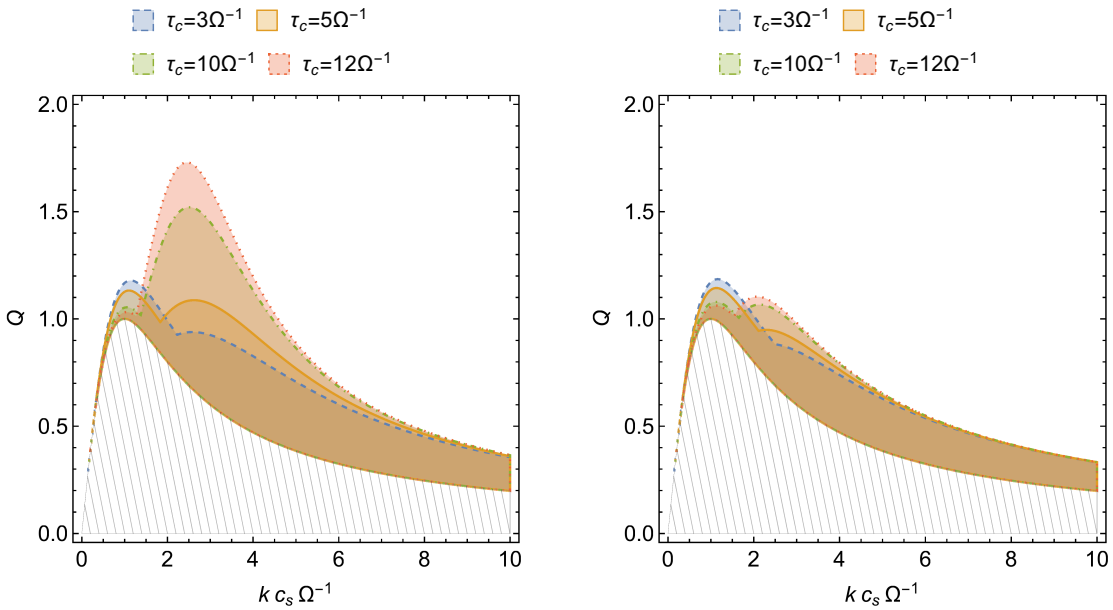
It is important to remember that the stability of the system is inherently linked to its properties. For instance, were one to remove the $+\gamma_s \frac{q\Omega}{k} A_h$ term from Equation 5.21, which originates from the kinematic viscosities not being functions of Σ,

the flow would be unstable to secular gravitational instability (Willerding, 1992; Gammie, 1996). This usually occurs in the long wavelength limit (i.e. $kc_s/\Omega \to 0$) when the flow is marginally stable to viscous instability (Equation 5.7); this is however beyond the scope of this work as the analysis presented here is unaffected by such instability.

The regions in the Q–kc_s/Ω plane where either (or both) of the Routh-Hurwitz stability criteria are not met, and the system is therefore unstable, are illustrated in Figure 5.7. The plot is obtained for $q = 3/2$, $\gamma = 5/3$, $\alpha_s = 0.05$, as well as $\mathrm{Pr} = 3$ (Fig. 5.7a) and $\mathrm{Pr} = 1$ (Fig. 5.7b); instability regions are obtained for various values of τ_c obeying the thermal balance, with the largest in each case being such that the disc is almost entirely viscously heated and therefore $e_{\mathrm{irr}} \simeq 0$. The unstable regions present two peaks, located at $kc_s/\Omega \sim 1$ and $kc_s/\Omega \sim 2.5 - 3$, respectively. The latter peak, which is usually dominant, is boosted by high Prandtl numbers (further highlighting the stabilising effect of thermal diffusion and cooling) and inefficient cooling, which means these factors also boost the overall instability; the sensitivity of the instability to the cooling time appears however to wane with a smaller Prandtl number. The peak located at $kv_s/\Omega \sim 1$ appears however insensitive to changes in the Prandtl number Pr, while fast cooling boosts it instead. Although the instability spans a substantial kc_s/Ω range, it prefers intermediate wavenumber values, meaning it is very relevant to zonal flow stability.

The results of a similar analysis are shown in Figure 5.8, with the difference of the adiabatic index being set to[2] $\gamma = 2$ this time. The larger value of γ used boosts the instability compared to the the $\gamma = 5/3$ case, particularly the peak located at $kc_s/\Omega \sim 2.5 - 3$, which in one case ($\mathrm{Pr} = 3$, $\tau_c = 8.5\Omega^{-1}$) reaches $Q \approx 2.7$; its sensitivity to the Prandtl number and cooling timescale remains however unaffected, with the instability being quenched by decreasing either value. Similarly the peak at $kc_s/\Omega \sim 1$ is still lightly boosted by quick cooling while mostly unaffected by varying the Prandtl number. The growth rates yielded by this method through Equation 5.58 present however a dependency on $1/\omega_0^2$, meaning that a singularity arises along the $\omega_0^2 = 0$ curve. It is therefore essential to test their accuracy by comparing them with the growth rates produced by the full dispersion relation. The testing was conducted for a similar set of parameters

[2]Although $\gamma = 2$ might not be the most physically relevant value of the adiabatic index, this has been constantly used in analyses of self-gravitating discs since the pioneering work by Gammie (2001). Its use here is therefore justified for comparing purposes with similar works in the literature.

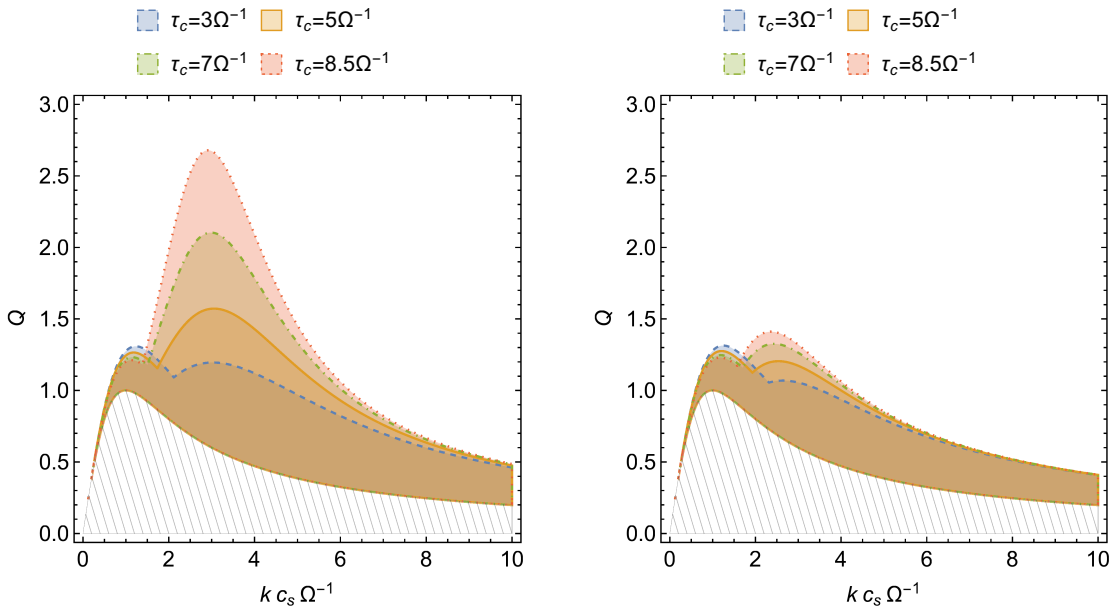


(a) Pr = 3, $\alpha_s = 0.05$ **(b)** Pr = 1, $\alpha_s = 0.05$

Figure 5.7 – Slow mode instability regions in the kc_s/Ω–Q space for $q = 3/2$, $\gamma = 5/3$, $\alpha_s = 0.05$ and Prandtl numbers **(a)** Pr = 3 and **(b)** Pr = 1. The regions are obtained for the cooling times $\tau_c = 3\Omega^{-1}$ (blue, dashed boundary), $\tau_c = 5\Omega^{-1}$ (orange, full), $\tau_c = 10\Omega^{-1}$ (green, dot-dashed) and $\tau_c = 12\Omega^{-1}$ (red, dotted). The unstable region features two peaks at $kc_s/\Omega \simeq 1$ and $kc_s/\Omega \simeq 2.5-3$. The latter peak is boosted by longer cooling times and larger Prandtl numbers, in which case the peak dominates over the other. The peak at $kc_s/\Omega \simeq 1$ appears instead insensitive to Pr and is slightly boosted by efficient cooling. The hatched region corresponds to where $\omega_0^2 < 0$ and the system is dynamically unstable.

as the longest cooling timescale in Figure 5.8a, with $\tau_c = 8.5\Omega^{-1}$, Pr = 3 and $kc_s/\Omega = 3$, and for a range of Q values obeying $0.7 \lesssim Q \lesssim 2.5$. While the accuracy of the chosen method's growth rates generally decreases with decreasing Q, the discrepancy between the two methods was found to be within 2% for most of the Q range, and within 5% for $Q \gtrsim 0.8$. This shows Equation 5.58 accurately represents the system's solutions for most of the Q range analysed.

The nature of the instability is analysed in Figure 5.9 for $\gamma = 2$, $\tau_c = 8.5\Omega^{-1}$ and both Pr = 3 and Pr = 1. As the instability can be triggered when either of the Routh-Hurwitz stability criteria (Equations 5.60, 5.61) is violated, the analysis considers the regions where each individual condition is not obeyed. Furthermore, the stability condition $c_1c_4 - c_2c_3 > 0$ (Eq. 5.61) can be further broken down

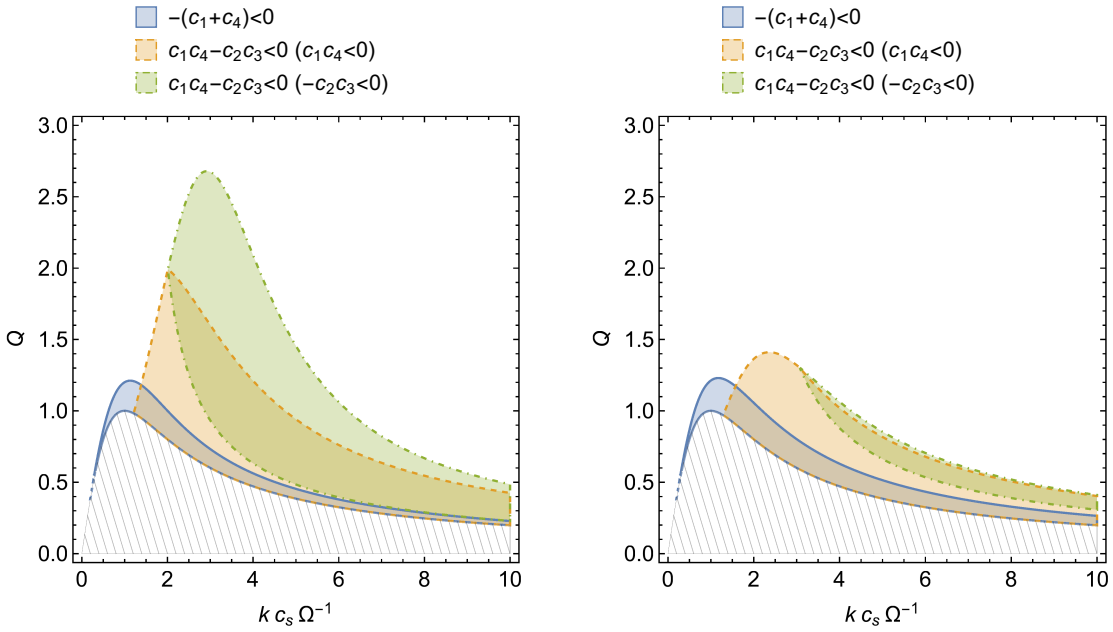


(a) Pr = 3, $\alpha_s = 0.05$ **(b)** Pr = 1, $\alpha_s = 0.05$

Figure 5.8 – Similar analysis of the slow mode unstable regions to Figure 5.7, but this time for an adiabatic index of $\gamma = 2$. The analysis is carried out for cooling times of $\tau_c = 3\Omega^{-1}$ (blue, dashed boundary), $\tau_c = 5\Omega^{-1}$ (orange, full), $\tau_c = 7\Omega^{-1}$ (green, dot-dashed) and $\tau_c = 8.5\Omega^{-1}$ (red, dotted); the values of the other parameters are unchanged from Figure 5.7. The larger value of γ boosts the instability, which reaches $Q \approx 2.7$ in one instance (Pr = 3, $\tau_c = 8.5\Omega^{-1}$). The hatched region again represents the are of the plane where $\omega_0^2 < 0$.

into its two components, where a negative contribution from $c_1 c_4$ would imply the instability is (at least partially) driven by growth in either PV or entropy, while $-c_2 c_3 < 0$ would mean their coupling is having a destabilising effect on the system. In the Pr = 3 case (Fig. 5.9a), the peak located at $kc_s/\Omega \sim 1$ appears to be caused by the $-(c_1 + c_4) > 0$ (blue, full boundary) stability criterion not being obeyed; this type of instability, as suggested by Equation 5.58, has an oscillatory nature. The nature of the peak can be further pinpointed as, assuming $\omega_0^2 > 0$, c_4 is always negative; this means the destabilising effect is produced by c_1, therefore by the entropy: this then corresponds to a thermal instability. The second peak, located at $kc_s/\Omega \sim 2.5 - 3$ is instead caused by the violation of the stability condition $c_1 c_4 - c_2 c_3 > 0$, meaning the instability here is non-oscillating. Both the action of entropy (orange, dashed boundary) and the PV-entropy coupling (green, dot-dashed) are found to be destabilising, with an overlap occurring where both

effects destabilise the system. Upon decreasing the Prandtl number to Pr = 1,



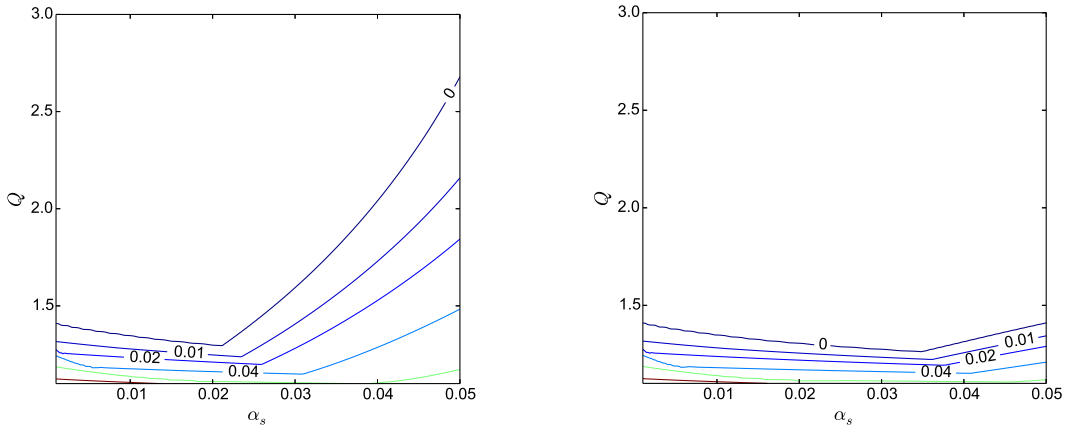
(a) Pr = 3 **(b)** Pr = 1

Figure 5.9 – Analysis into the nature of the slow mode unstable region, pinpointing which of the two stability criteria is violated in the unstable region obtained in Figure 5.8 for $\gamma = 2$, $\tau_c = 8.5\Omega^{-1}$ and Prandtl number **(a)** Pr = 3 and **(b)** Pr = 1. The peak at $kc_s/\Omega \sim 1$ is driven by the violation of $-(c_1 + c_4) > 0$ (blue, full boundary) which, given that $c_4 < 0$ for $\omega_0^2 > 0$, implies entropy is responsible for the instability, meaning this has a thermal nature. The peak at $kc_s/\Omega \sim 2.5 - 3$ is due to the $c_1c_4 - c_2c_3 > 0$ condition being violated, with both the c_1c_4 (again due to the entropy; orange, dashed) and the $-c_2c_3$ (due to PV-entropy coupling; green, dot-dashed) components being destabilising in different regions. Decreasing the Prandtl number quenches the destabilising effect of the PV-entropy coupling almost completely.

the destabilising contribution of the coupling is quickly suppressed, with it only contributing to a small part of unstable region while it was destabilising for a large part of the total unstable region in the Pr = 3 case. Moreover, the decrease in the Prandtl number caused a small boost to the unstable region caused by the violation of the $-(c_1 + c_4) > 0$ condition.

The remaining part of the analysis considers the effect of other parameters on the stability of slow modes; for this purpose, in each case the growth rates are optimised over k. Figure 5.10 illustrates how the growth rates depend on the dimensionless shear viscosity α_s and the Toomre parameter Q. The results

are obtained for $q = 3/2$, $\gamma = 2$, $\tau_c = 8.5\Omega^{-1}$ and a Prandtl number of Pr = 3 (Figure 5.10a) and Pr = 1 (Figure 5.10b). The whole range of α_s used is allowed by



(a) Pr = 3 **(b)** Pr = 1

Figure 5.10 – k-optimised slow mode growth rates as a function of the shear viscosity α_s and the Toomre parameter Q for $q = 3/2$, $\gamma = 2$, $\tau_c = 8.5\Omega^{-1}$ and a fixed Prandtl number of **(a)** Pr = 3 and **(b)** Pr = 1. The plots again show that a higher Prandtl number boosts the instability, but the value of α_s itself also appears to be a critical factor. For small α_s values ($\alpha_s \lesssim 0.02$), for which the disc is mostly heated by external irradiation, instability above $Q \simeq 1.4$ is not recorded regardless the value of Pr. As the shear viscosity value is increased, the two cases diverge with the Pr = 3 instance seeing the instability spread to $Q \simeq 2.7$ while no notable difference is observed in the Pr = 1 case.

thermal balance, with its value indicating the fraction of heat produced by viscous effects: small α_s imply the disc is mostly heated by an external irradiation, while for $\alpha \approx 0.05$ the heating occurs almost exclusively due to viscous effects. Although the Prandtl number still plays a key role in the stability of the system, it is clear from Figure 5.10 that the value of α_s is also critical to slow mode stability. In either case no instability is recorded for $Q \gtrsim 1.4$ when $\alpha_s \lesssim 0.02$, while for higher α_s values the two cases diverge: no significant difference is noticed for the Pr = 1 instance, while if Pr = 3 the system is unstable up to $Q \simeq 2.7$ for $\alpha_s = 0.05$. It is therefore apparent that the source of disc heating is crucial to disc stability, with a viscously heated disc being much more susceptible to slow mode instability.

Finally, the dependence of the k-maximised growth rates on Q and the cooling time τ_c is explored in Figure 5.11, with $\alpha_s = 0.05$ and Pr = 3. This again shows a

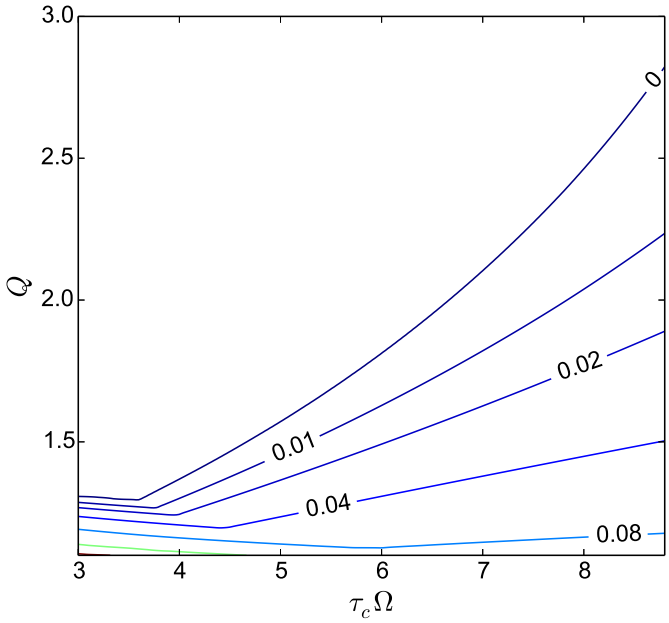


Figure 5.11 – Dependence of the k-optimised slow mode growth rate on the Toomre parameter Q and the cooling timescale τ_c. The analysis is carried out for $q = 3/2$, $\gamma = 2$, $\alpha_s = 0.05$ and $\mathrm{Pr} = 3$. In a large part of the Q range considered, shortening the cooling time has a stabilising effect as the peak located at $kc_s/\Omega \sim 2.5 - 3$ in Figures 5.7 and 5.8 is quenched by efficient cooling. The inverse however takes place for $Q \lesssim 1.25$ due to the peak located at $kc_s/\Omega \sim 1$ being instead lightly boosted by rapid cooling.

boosted instability for longer cooling times, which is in accordance with Figures 5.7 and 5.8; inefficient cooling times also correspond to a higher fraction of heat generated by viscous effects, which as seen in Figure 5.10 is crucial to boosting this instability. For most of the Q range considered, the instability is then quenched (partially or completely) by shortening the cooling time; this is due to the peak located at $kc_s/\Omega \sim 2.5 - 3$ in Figures 5.7 and 5.8, which is stabilised by rapid cooling. However the situation changes when $Q \lesssim 1.2$: here lengthening the cooling time has the opposite effect of boosting the instability. This is due to the peak at $kc_s/\Omega \sim 1$, having a thermal nature and being boosted by fast cooling, coming into play.

5.4 Summary

The work presented in this chapter consisted of a stability analysis of a viscous, compressible and self-gravitating Keplerian disc by means of a linear analytical calculation. Apart from the bulk and shear viscosities, the flow was also subject to a horizontal thermal diffusion and a constant cooling, with a freely-evolving axisymmetric structure present in the system. In the ideal case (no viscous/diffusive effects nor cooling) the system's four solutions – two density waves and two slow modes – were well known; however the introduction of the additional viscous, diffusive and cooling effects made the stability of the system non-trivial.

The stability analysis of the density wave modes was somewhat easier, as it was possible to use the regular perturbation method to obtain the individual contribution from each type of damping coefficient (bulk, shear and thermal) to the total growth rate; if their sum was positive, the system was unstable. The bulk viscosity and thermal diffusion were found to always create a stabilising effect to density wave modes for a Toomre parameter $Q > 1$, but the shear viscosity contribution was more complex. Under the influence of shear viscosity and thermal diffusion (the bulk viscosity contribution being ignored as constant throughout the parameter space), the system was found to be overstable across a wide range of wavelengths for $Q < 2$, with fast cooling quenching the unstable region. A non-SG overstability was also detected if the adiabatic index γ was smaller than the analytically derived critical value $\gamma \approx 1.305$; this however required slow cooling and long wavelength perturbations, with non-SG overstability observed for wavelengths exceeding $16H$ for $\gamma = 1.3$. The overstable growth rates, which had been maximised with respect to k, were then plotted as a function of adiabatic index and Toomre parameter for a fixed cooling time. This showed a gradual waning of overstable γ values as Q was increased, with only $\gamma \lesssim 1.25$ showing overstability for $Q \sim 5$, which was consistent with the analytically derived $\gamma \approx 1.305$ threshold.

On the other hand, the ideal solutions of the slow modes were degenerate, which meant that the regular perturbation method could not be applied in this case. Instead, equations for the temporal evolution of the zonal flow amplitude in potential vorticity and entropy were derived analytically, in such a way as to remove dependencies on zonal flow amplitudes in other quantities. An appropriate stability criterion (the Routh-Hurwitz criterion) was applied to impose stability conditions on the coefficients in the two equations. In the limits of short and long

wavelength perturbations, the system was found to be stable, in agreement with the classical approach. However, this work provided an extension to the classical stability analysis by considering perturbations of intermediate wavelength, where the system was found unstable to slow modes and the PV-entropy coupling to play an important role in the stability. The instability appeared boosted by high Prandtl numbers and higher adiabatic index values, as well as by stronger self-gravitating conditions (i.e. lower Q); also of importance was the source of disc heating, with viscous heating being more destabilising than external heating. Efficient cooling was however found to be detrimental to the instability, as long as $Q \gtrsim 1.25$.

This means that zonal flows could form and grow due to the presence of viscous and thermal effects in the right conditions, especially in a setup where heating is predominantly provided by viscous effects. As explored in Chapter 4, these zonal flows would however be potentially subject to disruptive non-axisymmetric instabilities in both self-gravitating and non-self-gravitating conditions. The two processes could theoretically form a sustained cycle, allowing an axisymmetric structure to exist and to influence the disc dynamics for an extended period of time.

Chapter 6

Self-sustenance of gravitational turbulence

6.1 Introduction

As previously mentioned in Section 2.2.2, a self-gravitating disc can enter a turbulent thermally-regulated state if its cooling is not too efficient, with Gammie (2001) finding such a state is achieved if $\tau_c \Omega = \beta > 3$. In this case the Toomre parameter of the disc, representing the (inverse) strength of self-gravity, would oscillate around $Q_{\rm crit} \sim 1$. While this gravito-turbulent state can be of importance in the transport of angular momentum towards the central object, therefore allowing accretion to take place, the details of its self-sustenance are unclear. If the Toomre parameter of the disc exceeds unity, the shearing sheet model used throughout this thesis only allows a transient amplification of non-axisymmetric shearing waves. In the linear regime these transient growths are however quickly quenched by damping viscous effects, meaning the system is linearly stable to non-axisymmetric disturbances. These transiently growing shearing waves can however potentially be continuously regenerated in the presence of a coupling with a non-linear feedback, although the dynamics of this process are not well understood. This coupling then enables energy to be continuously extracted from the background flow and fed into the non-axisymmetric gravitational instability, ensuring the sustenance of the gravito-turbulent state.

As mentioned in Section 2.3, zonal flows have been observed in hydrodynamical and MHD simulations analysing turbulent states of different origins, so it is possible

that zonal flows may be involved in the quasi-steady, gravito-turbulent regime. If that is the case, it would be important to understand whether the presence of zonal flows plays a key role in the self-sustenance of turbulence itself, and address how this potentially occurs. Were that to be the case, the axisymmetric structure of the zonal flow would need to be periodically regenerated in order to allow the turbulent state to be maintained; this can be achieved in two ways, by means of an axisymmetric instability or by the non-linear interaction of two non-axisymmetric waves such that the azimuthal components of their wavenumbers cancel. The turbulent feedback is therefore likely to include the triggering of an instability, but it is unclear in which variables this would manifest. Potential vorticity and entropy have been core quantities in the works presented in Chapters 4 and 5, so it is possible they also play a role in this case.

6.2 Initial conditions

Following the example set by several other accretion disc simulations (e.g. Johnson and Gammie, 2005; Shen et al., 2006), the runs in `CASPER` were initialised with a random white noise perturbation in both radial and azimuthal velocities u and v within the Fourier space range of wavenumbers satisfying

$$k_{\min}^2 < k^2 < k_{\max}^2, \tag{6.1}$$

where the limiting wavenumber values are set as

$$k_{\min} = \frac{2\pi}{L}, \qquad\qquad k_{\max} = 32 k_{\min}. \tag{6.2}$$

The white noise employed possessed a uniform distribution in the range $[-0.5, 0.5]$ on which a $10^{-3} c_s$ scaling factor is applied. Both velocity components were set to zero outside of this 2-dimensional range. These random initial conditions were then transformed to real space by means of Fourier transforms using the `FFTW` routine.

Random perturbations were however not applied to surface density and internal energy, with their initial values (applied in real space directly) being uniform and given by

$$h = \ln \Sigma_0 = 0, \tag{6.3}$$

$$e = e_0 = \frac{c_s^2}{\gamma(\gamma - 1)}. \tag{6.4}$$

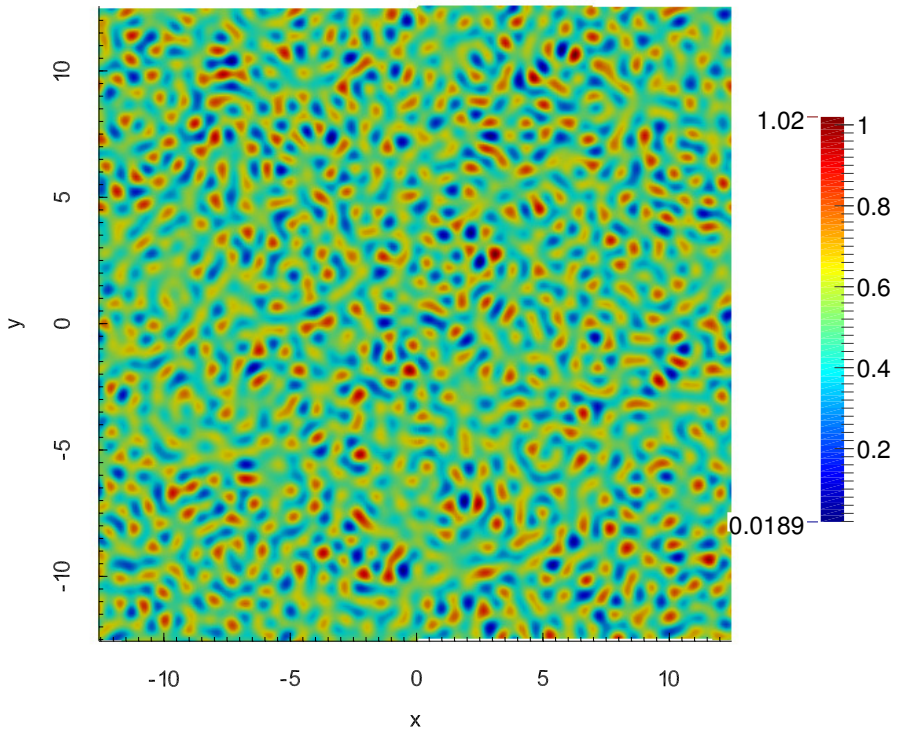


Figure 6.1 – Initial conditions applied to the potential vorticity visualised in real space with $L_x = L_y = 8\pi\,(\pi G \Sigma_0)/\Omega^2$. This corresponds to white noise initial conditions in both velocity components for $k^2_{\rm min} < k^2 < k^2_{\rm max}$.

Various simulations were run, with initial values of Q in the range of $1 \leq Q_0 \leq 2$, a constant cooling in the $7 \leq \tau_c\,\Omega \leq 15$ range, an adiabatic index $\gamma = 2$ (after Gammie (2001) and enabling a direct comparison with much of the literature) and possessing varying spatial resolutions. The high resolution runs (1024×1024 grid points) were all conducted with a box size of $L = 8\pi\,(\pi G \Sigma_0)/\Omega^2$ (where the mean surface density Σ_0 is conserved), as this allows the viscosities and thermal diffusion

to remain at the initialised constant value ($\alpha_s \sim 0.004$, instead of being adapted by the numerical method to ensure numerical stability), while also possessing a sufficient magnitude to permit dynamical stability of the flow. This choice of box size may of course somewhat impact the development of the gravito-turbulent state, as the auto-correlation function in Gammie (2001) shows it requires a box that is elongated in the y direction to fully resolve the turbulence. Larger box scenarios with the same 1024×1024 resolution are however prohibitive in the current code configuration due to the computational time required; the behaviour of the turbulent state in larger boxes was investigated in lower resolution runs, with no qualitative difference noticed.

Figure 6.1 shows the initial conditions in the potential vorticity in the (x, y) space; this presents a random, chaotic character due to its dependency on the azimuthal velocity (to which a random white noise is applied). This random behaviour in the velocity components and related quantities was employed as a way of mimicking the conditions of residual turbulence which take place in the early life of a disc following the collapse of the parent molecular cloud (e.g. Godon and Livio, 2000). It is also important to notice that no incompressibility restriction was applied to the flow (unlike in the works by Johnson and Gammie, 2005; Shen et al., 2006), with the velocity components therefore not being correlated.

6.3 Non-linear evolution

Applying the initial conditions outlined above, as well as a constant radiative cooling, viscosities and thermal diffusion (as outlined in Chapter 3), the system was allowed to evolve for several tens of orbits.

Figure 6.2 illustrates the temporal evolution of the average Toomre parameter $\overline{Q} = \sqrt{\gamma(\gamma-1)\overline{e}}$, where $\bar{e} = \tilde{e}(k_x = 0, k_y = 0)$ is the mean internal energy. Due to the ICs not being in thermal equilibrium (as discussed more in detail later), the disc initially cools towards an equilibrium temperature; this is however prevented by the gravitational instability being triggered at $t\Omega \simeq 7$. In these strongly gravitating conditions, with $\overline{Q} \sim 1$, the disc quickly heats up due to viscous effects and it subsequently enters a turbulent state from $t\Omega \gtrsim 50$. In this gravito-turbulent regime, characterised by an average $\overline{Q}$ value of $\overline{Q} \sim 2$, a recurrence of heating events with a characteristic timescale of $\sim 50 - 60\Omega^{-1}$ (i.e. $8 - 10$ orbits) takes place. A short-timescale modulation can also be observed in $\overline{Q}$, with its origin

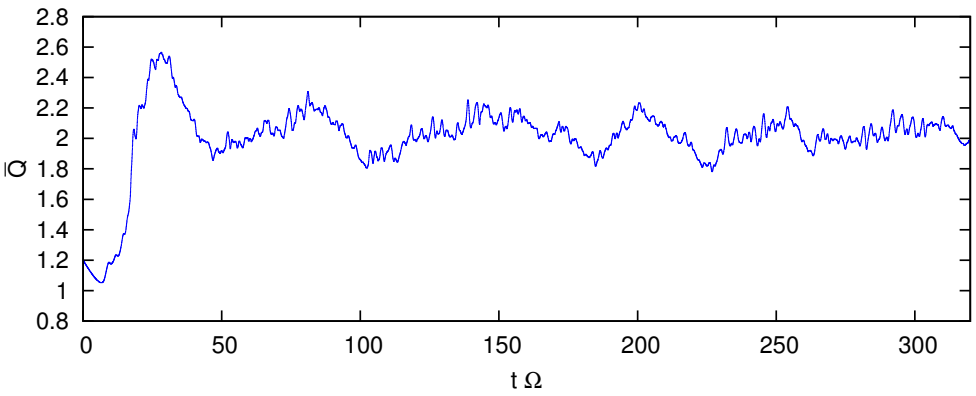


Figure 6.2 – Temporal evolution of the average value of the Toomre parameter Q. After a burst of strong GI, the disc settles into a state of gravito-turbulence with an average Q value of $\overline{Q} \sim 2$.

being investigated later in the chapter.

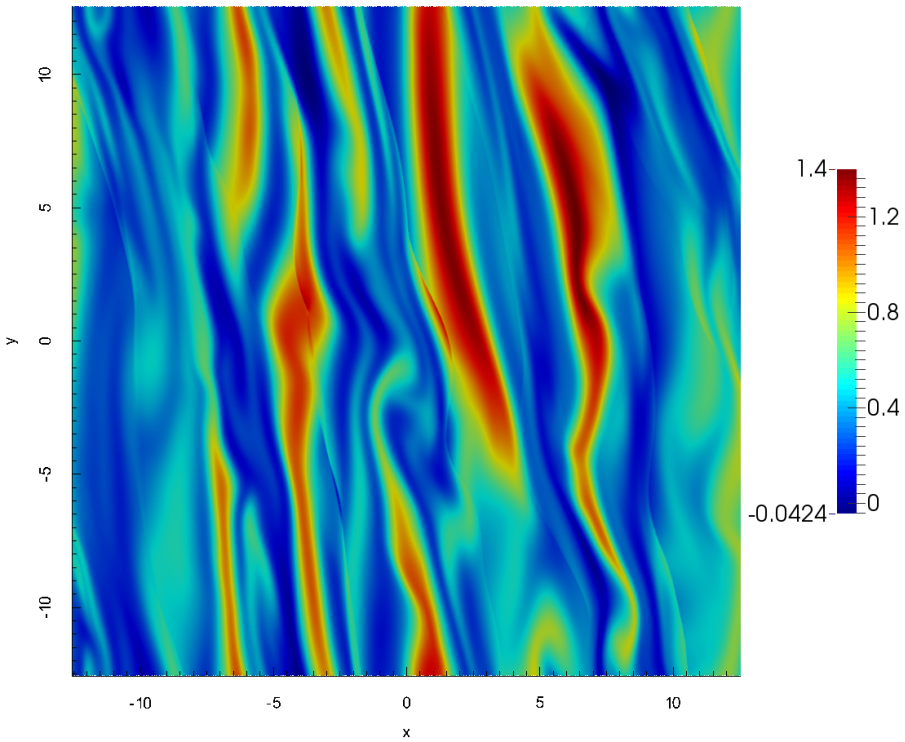


Figure 6.3 – Illustration of the nearly axisymmetric structure visible in the entropy. A similar structure is also observed in other quantities, such as PV and density. This structure is believed to play a part in the self-sustenance of the turbulent state.

Having obtained a state of gravito-turbulence, it was important to address the means by which it managed to self-sustain. The gravito-turbulence appeared to be accompanied by the presence of a nearly axisymmetric structure in the specific entropy and other quantities (such as potential vorticity and surface density); an example of this is shown in Figure 6.3. The structure illustrated appears to be localised to a portion of the sheet rather than encompassing its entirety; this is believed to be simply owed to the difficulty of maintaining such an extended structure, with parts of it not causally connected in a supersonic shear flow.

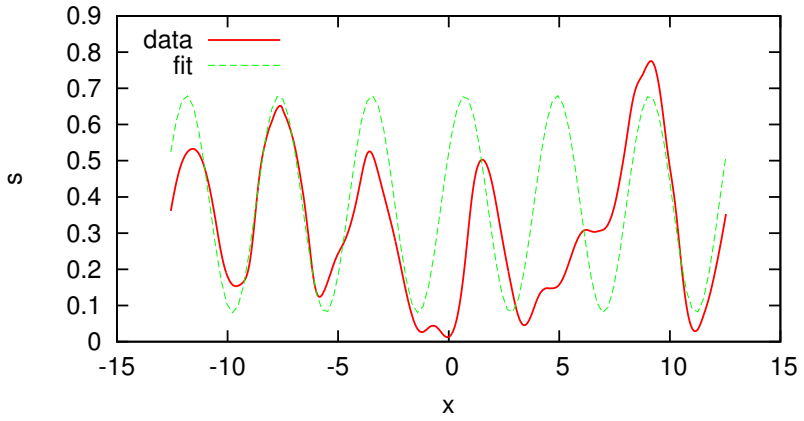


Figure 6.4 – Fit of the y-averaged specific entropy (red, full line) of a snapshot similar to Figure 6.3 with a sinusoidal function (green, dashed) to infer the wavenumber of the axisymmetric structure (in this case $k_x \pi G \Sigma_0 / \Omega^2 = 1.25$) present in the flow.

The wavelength of the observed axisymmetric structure can be estimated by taking the y-average of real space frames showing the presence of the zonal flow. An example of the y-averaged profiles obtained is shown in Figure 6.4, where the resulting radial profile in the specific entropy is fitted with a sinusoidal function. The fit in this case has a wavenumber of $k_x \pi G \Sigma_0 / \Omega^2 = 1.25$.

Further analysis of the simulation output showed that two axisymmetric structures with adjacent wavenumbers were active in the flow, as shown in the power spectrum time maps of PV and entropy in Figure 6.5. This shows that the zonal flows possessing the wavenumbers in question are enhanced during heating events in the Toomre parameter and decay as the disc begins to cool afterwards, therefore suggesting zonal flows might be playing an important role in the self-sustenance of the gravito-turbulent state. The wavenumbers of the two dominant zonal flows, as shown in Figure 6.5, are not fixed at $k_x \pi G \Sigma_0 / \Omega^2 = 1$ and 1.25 at all times, al-

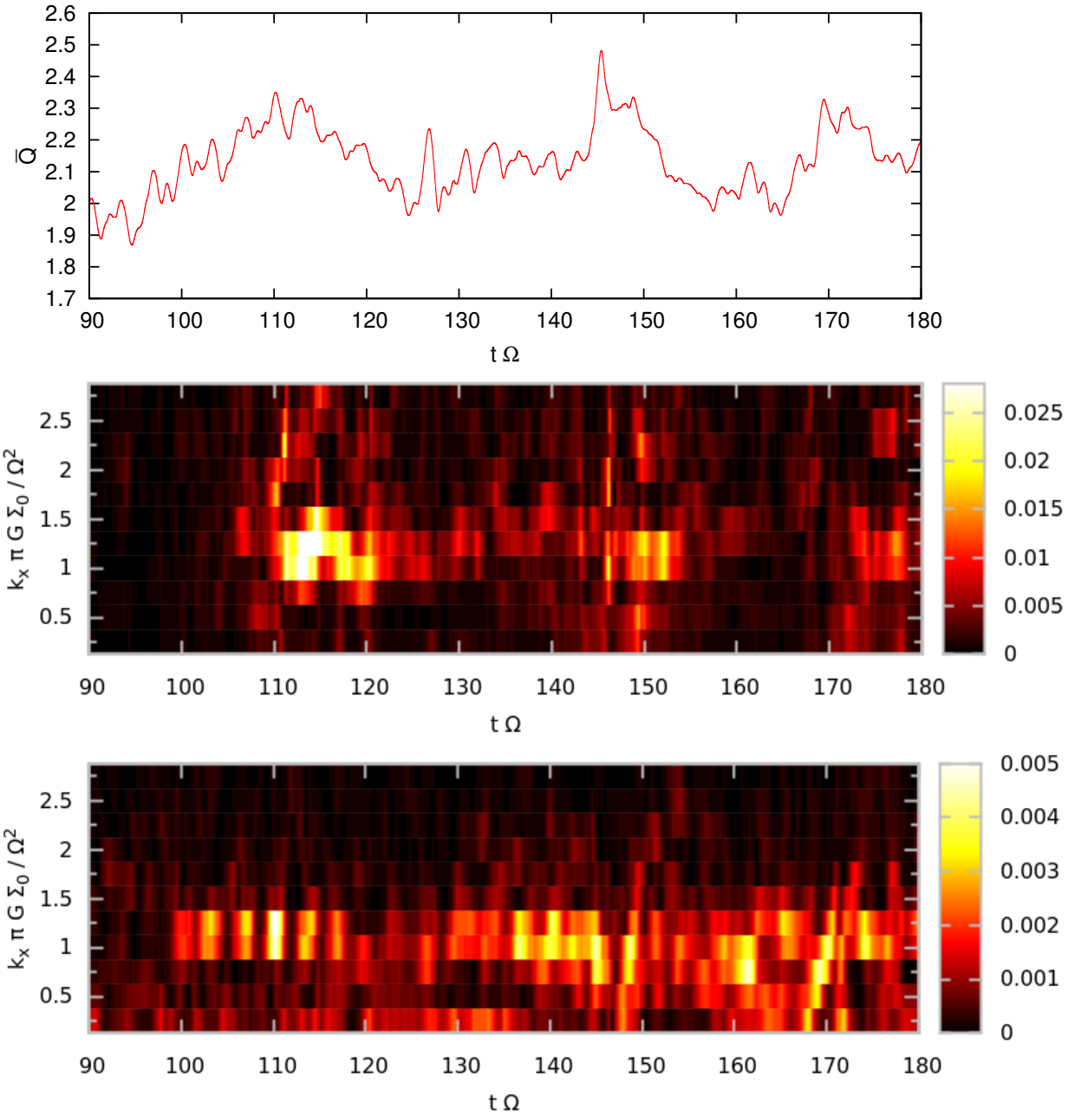


Figure 6.5 – Interval between $t\Omega = 90$ and $t\Omega = 180$ for $\tau_c = 10\Omega^{-1}$ showing the average Toomre parameter $\overline{Q}$ (***top***) and the $k_y = 0$ power spectrum maps for $0.25 \leq k_x \pi G \Sigma_0 / \Omega^2 \leq 2.75$ for PV (***middle***) and entropy (***bottom***). This shows the wavenumbers $k_x \pi G \Sigma_0 / \Omega^2 = 1.00$ and 1.25 dominating over the other values, especially during heating events in the flow.

though the dominant wavenumbers do not usually stray far from $k_x \pi G \Sigma_0 / \Omega^2 \sim 1$. Furthermore, lower resolution runs experimenting with varying box sizes and resolutions found no significant difference in the wavenumbers of the dominant zonal flows, suggesting that these only depend on the stability properties of the flow, as discussed in Section 6.4 together with the issue of structure regeneration.

6.3.1 Reynolds and gravitational stresses

An important part of the disc dynamics are the components of the flow's stress tensor; in particular, the Reynolds (or hydrodynamical) and gravitational stresses can be used to calculate the amount of angular momentum transported through the shearing sheet and to estimate how much of this is caused by gravitational effects.

The spatially-averaged gravitational stress $\langle G_{xy} \rangle$ can be easily calculated in Fourier space making use of Parseval's theorem (which is derived and verified in Appendix C) according to

$$\langle G_{xy} \rangle = \frac{1}{L_x L_y} \int_0^{L_x} \int_0^{L_y} G_{xy} \, \mathrm{d}x \, \mathrm{d}y = -\frac{1}{4\pi G} \sum_{\boldsymbol{k}} \tilde{g}_x(\boldsymbol{k}) \tilde{g}_y(\boldsymbol{k})^*, \tag{6.5}$$

where

$$G_{xy} = -\frac{g_x g_y}{4\pi G} \tag{6.6}$$

is the local stress and

$$\tilde{g}_i = -\mathrm{i} k_i \left| \tilde{\Phi}_d(\boldsymbol{k}) \right|. \tag{6.7}$$

This results in

$$\langle G_{xy} \rangle = -\frac{1}{4\pi G} \sum_{\boldsymbol{k}} k_x k_y \left| \tilde{\Phi}_d(\boldsymbol{k}) \right|^2. \tag{6.8}$$

However the expression for the disc potential (Equation 3.8) also describes the behaviour of self-gravity away from the disc's midplane, which is described by $\left| \tilde{\Phi}_d(\boldsymbol{k}) \right| \propto \mathrm{e}^{-|\boldsymbol{k}||z|}$. Vertical integration of G_{xy} therefore yields

$$\int G_{xy} \mathrm{d}z \propto \int_{-\infty}^{+\infty} \left| \tilde{\Phi}_d \right|^2 \mathrm{d}z \propto 2 \int_0^{+\infty} \mathrm{e}^{-2|\boldsymbol{k}||z|} \mathrm{d}z = \frac{1}{|\boldsymbol{k}|}. \tag{6.9}$$

The extra factor of $1/|\boldsymbol{k}|$ therefore alters the complete form of the spatially-averaged gravitational stress to

$$\begin{aligned}\langle G_{xy}\rangle &= -\frac{1}{4\pi G}\sum_{\boldsymbol{k}}\frac{k_x k_y}{|\boldsymbol{k}|}\left|\tilde{\Phi}_{\mathrm{d,m}}(\boldsymbol{k})\right|^2 \\ &= -\pi G\sum_{\boldsymbol{k}}\frac{k_x k_y}{|\boldsymbol{k}|^3}\left|\tilde{\Sigma}(\boldsymbol{k})\right|^2 \mathrm{e}^{-2s|k|}. \end{aligned} \tag{6.10}$$

The spatially-averaged Reynolds stress $\langle H_{xy}\rangle$, on the other hand, is calculated in real space in order to avoid a Fourier space convolution, with its expression being

$$\langle H_{xy}\rangle = \frac{1}{L_x L_y}\int\int H_{xy}\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{L_x L_y}\int\int \Sigma uv\,\mathrm{d}x\,\mathrm{d}y, \tag{6.11}$$

where H_{xy} is the local Reynolds stress.

When analysing both stresses, it was noticed that the Reynolds stress possessed several troughs with $\langle H_{xy}\rangle < 0$; negative Reynolds stress values are usually associated with leading density waves, as these add energy to the background flow. The evidence regarding the correlation between negative Reynolds stress and leading power was however inconclusive. Moreover, the modulation on $\langle H_{xy}\rangle$ appeared more irregular than that observed in other quantities, suggesting secondary effects such as epicyclic oscillations may be at play.

Moving window box-car averaging

Both stresses were then smoothed out by time-averaging during the simulations using the moving window box-car technique, with the size of the window (i.e. averaging period) being around $1/3-1/2$ of an orbit, leading to the smoothing down of only short-period irregular modulations. An example of the original ($\langle H_{xy}\rangle$) and time-averaged ($\langle H_{xy}^{\mathrm{avg}}\rangle$) Reynolds stresses is shown in Figure 6.6.

The time-averaged stress (red, full line) retains the larger amplitude modulation occurring during heating events (e.g. for $t\Omega \simeq 341$), while the irregular oscillations taking place in the original stress (blue, dotted) between heating events are smoothed out.

The time-averaged stresses are then used to calculate the rate of angular momentum transport through the disc by means of

$$\alpha = \frac{\langle G_{xy}^{\mathrm{avg}} + H_{xy}^{\mathrm{avg}}\rangle}{qP}, \tag{6.12}$$

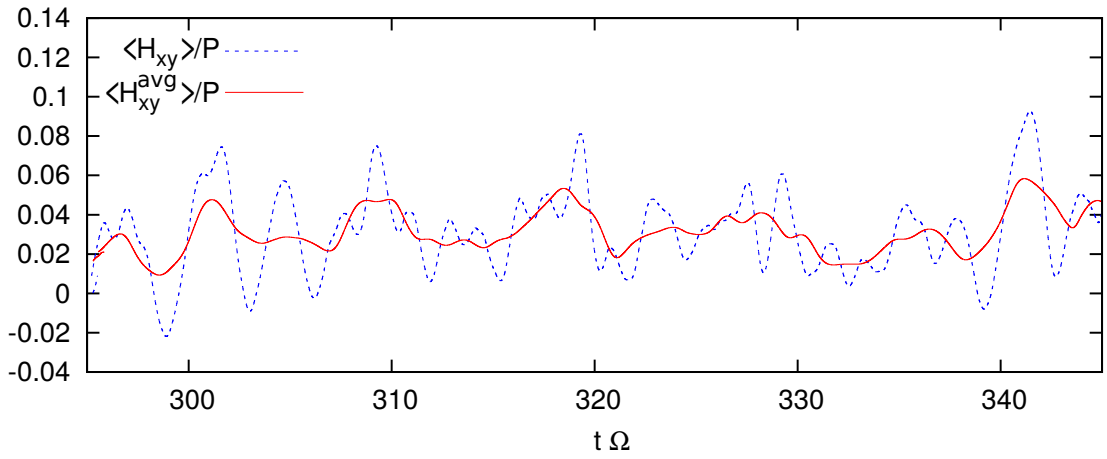


Figure 6.6 – Temporal evolution of both measured ($\langle H_{xy} \rangle$, blue dotted line) and time averaged ($\langle H_{xy}^{\mathrm{avg}} \rangle$, red full line) Reynolds stress, with both stresses being divided by the total pressure.

which can be rewritten as

$$\alpha = \alpha_{\mathrm{grav}} + \alpha_{\mathrm{Reyn}}. \tag{6.13}$$

Figure 6.7 illustrates the temporal development of both components of α, as well

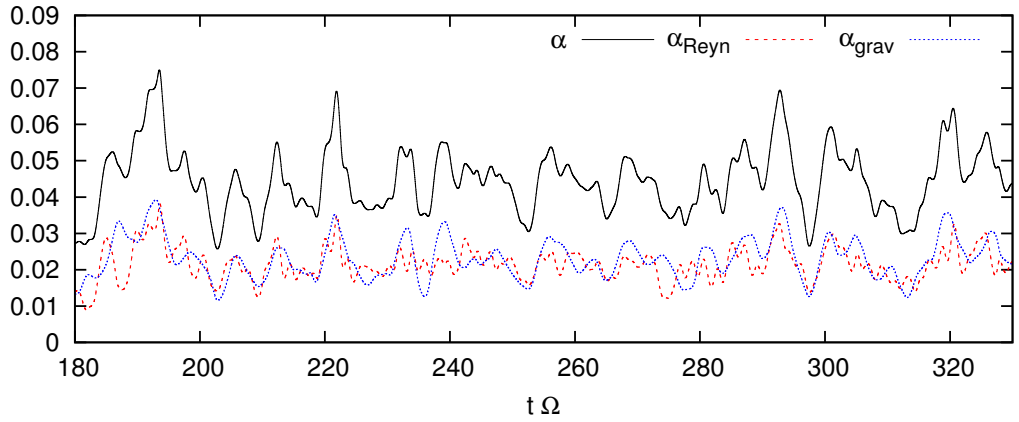


Figure 6.7 – Temporal evolution of the α parameter (black, full line), as well as its hydrodynamical (red, dashed) and gravitational (dotted, blue) components. The total value of α averages around $\alpha \sim 0.044$, with the Reynolds and gravitational components each providing roughly half of the angular momentum transport.

as their sum. The average value of the transport rate of angular momentum lies around $\alpha \sim 0.044$. This is comparable with the α values recorded by similar self-

gravitating simulations (e.g. Gammie, 2001; Mamatsashvili, 2011). The hydrodynamical ($\alpha_{\rm Reyn}$, red dashed line) and gravitational ($\alpha_{\rm grav}$, blue dotted) components possess similar average values, with each therefore contributing around half of the angular momentum transport; this is again consistent with similar simulations.

6.3.2 Turbulent viscosities

The non-linear simulations carried out in `CASPER`, unlike the axisymmetric linear analysis described in Chapter 5, did not take into account the internal energy $e_{\rm irr}$ to which the disc would relax in the absence of viscous heating due to the presence of a source of external irradiation. The consequence of considering a disc with $e_{\rm irr} = 0$ was that the thermal balance equation (Equation 5.12) was changed to

$$\alpha_s \tau_c = \frac{1}{q^2 \Omega (\gamma - 1)}, \tag{6.14}$$

meaning for a given value of τ_c there exists only one value of α_s satisfying the equation, and vice versa. The runs were initialised with a viscosity of order $\alpha_s \sim 0.004$ which would only be thermally balanced by a cooling time ($\tau_c \sim 100\Omega^{-1}$) that is roughly one order of magnitude longer than those used.

The temporal evolution of the average Toomre parameter $\overline{Q}$ shown in Figure 6.2 however clearly illustrates that the system has reached a quasi-steady state where the temperature (i.e. $\overline{Q}$) remains roughly constant. As the initialised viscosities remained within 10% of their initial values at all times, it is clear that thermal balance could only be achieved thanks to a viscous enhancement caused by turbulent motions.

It is straightforward to calculate the expected effective shear viscosity $\alpha_{\rm eff} = \alpha_{\rm turb} + \alpha_{\rm init}$ allowing thermal balance, as this can be found by solving Equation 6.14 for the appropriate cooling time. It is also possible to compare the theoretical shear viscosity needed for thermal equilibrium with the sum of the simulations' turbulent (i.e. from the Reynolds and gravitational stresses, by means of Equation 6.12) and initial shear viscosities. The results of this comparison are shown in Figure 6.8 where the average of the total observed shear viscosity $\alpha_{\rm eff}$ is plotted as a function of the inverse of the imposed cooling timescale. The plot shows that the observed average effective viscosities (black data points) lie close to their respective expected values given by Equation 6.14 (dashed line), all observed values falling within 10% of the theoretical effective viscosity (shaded area).

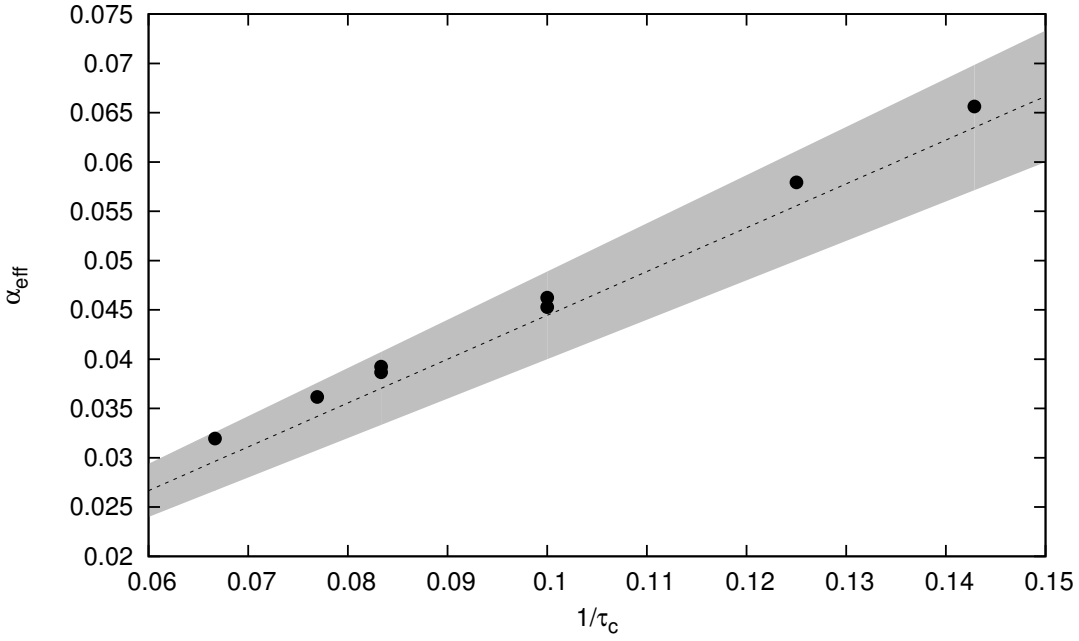


Figure 6.8 – Linear relationship between α_{eff} and τ_c^{-1}, as suggested by Equation 6.14 for a few values of the cooling timescale. The dashed line represents Equation 6.14 itself, while the shaded area indicates the values of α_{eff} falling within 10% of the dashed line for a given τ_c^{-1}. Some $1/\tau_c$ values have multiple data points, but these appear almost indistinguishable from one another due to the similarity between their average values of α_{eff}.

If there exists an enhancement to the shear viscosity due to turbulent motions, a similar turbulent increase must also occur for the thermal diffusion. This is however much harder to estimate due to the absence of a net heat flux across the system (arising from the lack of a temperature gradient), and it would require detailed heat transport analysis which `CASPER` was not designed to compute. Instead, the effective thermal diffusion is estimated through the turbulent Prandtl number Pr_{turb}. The inability of gravitational interactions to transport heat and the assumption that fluid motions carry heat in a similar way in which they transport angular momentum leads to the following approximation for the turbulent Prandtl number:

$$\text{Pr}_{\text{turb}} = \frac{\nu_{s,\text{turb}}}{\nu_{t,\text{turb}}} \sim \frac{H_{xy} + G_{xy}}{H_{xy}}. \tag{6.15}$$

Figure 6.9 shows an example of the estimation of the turbulent Prandtl number

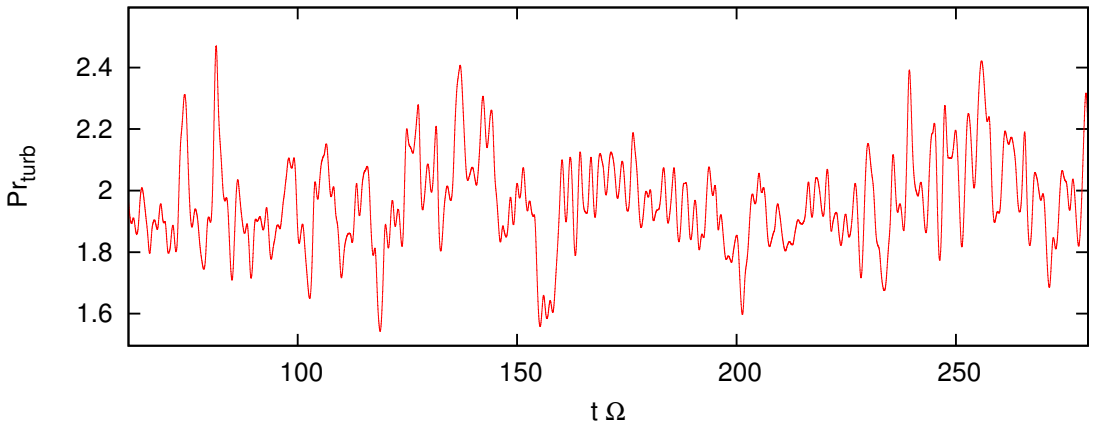


Figure 6.9 – Estimation of the turbulent Prandtl number using the xy stress. The average Prandtl value is $\mathrm{Pr}_{\mathrm{turb}} \simeq 1.9$, though it oscillates somewhat either way.

using Equation 6.15, with its average value being $\mathrm{Pr}_{\mathrm{turb}} \simeq 1.9$. Other runs feature an average turbulent Prandtl number of roughly 2, with oscillations that can extend from $\mathrm{Pr}_{\mathrm{turb}} \simeq 1.3$ to roughly 3.

6.4 Instabilities and self-sustenance

6.4.1 Axisymmetric instability

As seen in Figure 6.5, the dominant axisymmetric structures do not gradually decay with time, but are instead continuously regenerated throughout the simulation. This appears to suggest that an instability is allowing the zonal flows to grow transiently during heating events; the zonal flows subsequently decay as the instability criterion is no longer fulfilled while the disc cools back down.

One plausible hypothesis is that the zonal flows are growing as a result of the slow mode instability encountered in Chapter 5, the stability properties of which are discussed in Section 5.3.3. It is important to remember that various quantities affect the shape of the unstable region in the kc_s/Ω–Q plane; these include: cooling timescale, adiabatic index, turbulent Prandtl number and the fraction of heat generated by viscosity. For a given run the former two quantities are kept fixed and therefore do not influence the stability properties of the slow mode instability. The heat fraction produced by viscosity is likewise fixed at $f_{\mathrm{visc}} = 1$, due to the absence of an external irradiation term in the internal energy equation,

but the value of the effective viscosity does however evolve with time as the system oscillates around an average state of thermal balance, affecting the shape of the instability region. The Prandtl number does likewise exhibit a temporal evolution within a run.

Figure 6.10 illustrates the temporal evolution of the relevant parameters (Q, $\alpha_{\rm eff}$ and $\mathrm{Pr}_{\rm turb}$), and the resulting slow mode instability region in the kc_s/Ω–Q plane for six time snapshots (**a**–**f**) for a run with $\tau_c = 13\Omega^{-1}$; the instability regions also feature data points for the wavenumbers (converted into acoustic units) of the dominant axisymmetric structures at that particular time.

When the system reaches the Q peak of a weak heating event in (**a**) the value of the effective shear viscosity $\alpha_{\rm eff}$ is already dropping due to the thermally-regulated feedback between viscous heating and radiative cooling; the diminished value of $\alpha_{\rm eff}$, together with the turbulent Prandtl number hitting a trough, cause the system to be stable. While the disc cools down on its way to reach snapshot (**b**), $\alpha_{\rm eff}$ starts to increase which means that the disc – also thanks to the slightly higher value of $\mathrm{Pr}_{\rm turb}$ – is strongly unstable to slow modes at point (**b**). Subsequent heating causes another drop in the effective viscosity such that, although the Prandtl number retains a similar value to the previous snapshot, the system is stable at (**c**). A substantial hike in $\alpha_{\rm eff}$ (where it reaches ~ 1.7 times its average value) after a period of cooling triggers the instability again at (**d**), with the dominant wavenumbers slowly slipping towards less unstable conditions at (**e**) as Q increases and the size of the unstable area shrinks due to a drop in the viscosity. Lastly as the heating event reaches its peak at (**f**), a reduction in both $\alpha_{\rm eff}$ and $\mathrm{Pr}_{\rm turb}$ causes flow stability and allows the system to cool.

This analysis also explains why axisymmetric structures with specific wavenumbers are preferred at any one time: this is due to the shifting shape of the slow mode instability region, with the two dominant wavenumbers usually being centred around the peak of the region, where the growth rate is larger. It is therefore clear that the quantisation of wavenumbers originating from the consideration of a finite box will often result in the dominance of two axisymmetric structures due to the impossibility of a wavenumber coinciding with the region's peak for a period of time (as the wavenumbers also need to be multiplied by the ever-evolving Q to be converted into acoustic units).

It is important to remember that the analysis features some approximations, such as assuming that the turbulent viscosity can be modelled in the same way

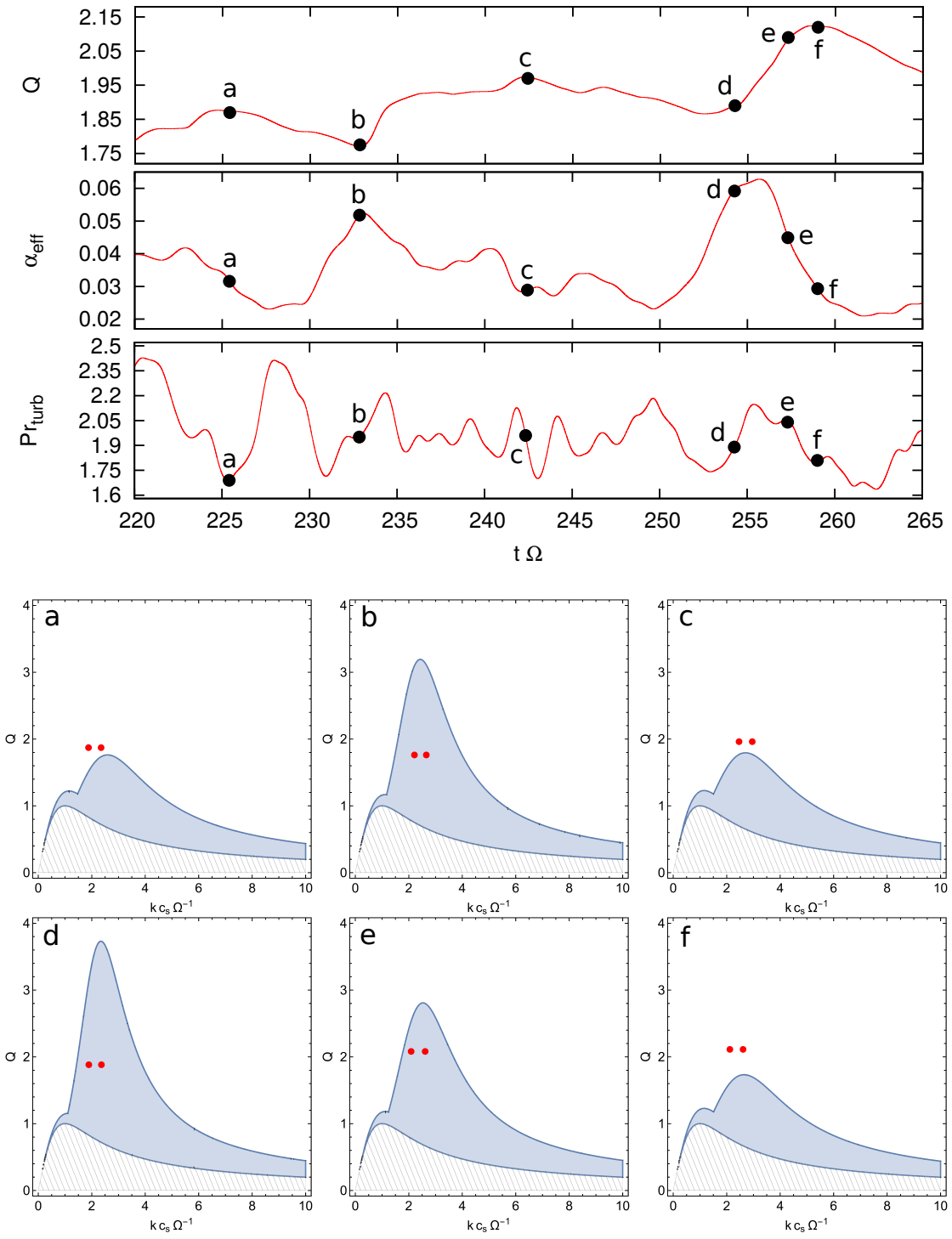


Figure 6.10 – Temporal evolution of Q, $\alpha_{\rm eff}$ and $\mathrm{Pr}_{\rm turb}$ (***Top***) and the position of the dominant axisymmetric structures (red data points) within the slow mode instability region at six time snapshots (***Bottom***) for a run with $\tau_c = 13\Omega^{-1}$.

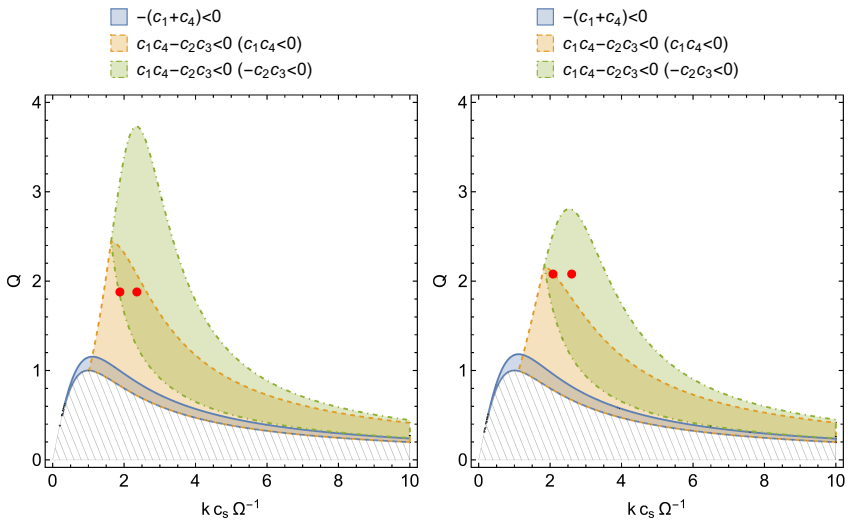


Figure 6.11 – Nature of the slow mode instability in frames (**d**) and (**e**) of Figure 6.10. The blue, full border regions indicate areas where the instability is driven by entropy and oscillatory in nature; the orange, dashed areas represent instances where it is again driven by entropy, but has a non-oscillatory behaviour; while the PV-entropy coupling provides a destabilising contribution in the green dot-dashed regions, the instability again possessing a non-oscillatory nature.

as the laminar disc viscosity (Balbus and Papaloizou, 1999), the estimation of the Prandtl number, its variability and the value of α_{eff} not being constant. Nevertheless, the results obtained from Figure 6.10 appear to strongly back up the idea that the slow mode axisymmetric instability plays a crucial and direct role in the self-sustenance of the gravito-turbulent state by providing a source of regeneration for the axisymmetric structures, as well as constraining their dominant wavelengths.

By means of the same technique used in Chapter 5, it was possible to pinpoint the nature of the slow mode instability acting on the flow by splitting its instability criteria into components. This is shown in Figure 6.11 for frames (**d**) and (**e**), respectively. The instability (whose criteria are given in Equations 5.60–5.61 in Section 5.3.3) may be due to entropy and be oscillatory ($-(c_1 + c_4) < 0$, blue full border) or non-oscillatory ($c_1c_4 - c_2c_3 < 0$ ($c_1c_4 < 0$), orange dashed) in nature, or may be non-oscillatory and caused by the PV-entropy coupling ($c_1c_4 - c_2c_3 < 0$ ($-c_2c_3 < 0$), green, dot-dashed). This analysis shows that in frame (**d**), when the system is in its most unstable configuration, the instability is driven by both entropy and by the PV-entropy coupling. As the system becomes less unstable in frame (**e**) the instability is instead mostly provided by the PV-entropy coupling.

6.4.2 Non-axisymmetric instability

The work carried out in Chapter 4 however shows that zonal flows of intermediate wavelength such as the ones generating here are subject to the disruptive effect of non-axisymmetric instabilities. A first hint this may be happening can be deduced from the axisymmetric power spectrum maps in Figure 6.5 – in particular that of the entropy – where, although the zonal flows are forced to grow by the axisymmetric instability during periods of heating in the disc (as seen in Figure 6.10), their amplitudes feature some irregular dips rather than growing exponentially. In order to investigate the cause of this behaviour, a k_y power spectrum map is constructed to assess whether any non-axisymmetric instability is also affecting the system; an example of this is shown in Figure 6.12 in both PV (top) and specific entropy (bottom).

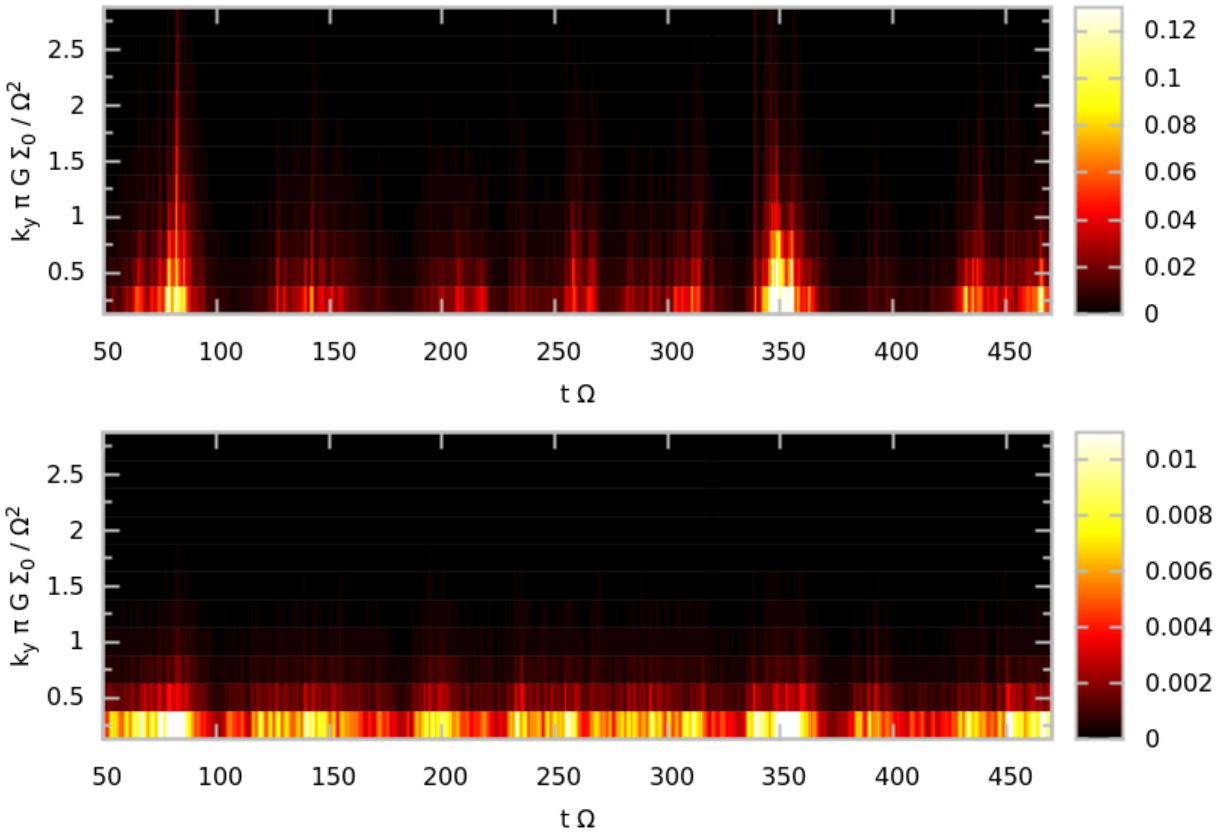


Figure 6.12 – k_y power spectrum maps for all k_x in PV (***top***) and entropy (***bottom***) for a run with $\tau_c = 12\Omega^{-1}$. This illustrates the presence of a non-axisymmetric structure with wavelength $k_y \pi G \Sigma_0 / \Omega^2 = 0.25$ in both quantities.

The plot illustrates the presence of a non-axisymmetric structure, plausibly related to an instability, in the entropy and PV which possesses a wavelength matching the box size (i.e. $k_y \pi G \Sigma_0 / \Omega^2 = 0.25$). In order to better understand the role played by the discovered non-axisymmetric instability, a more basic test run was set up where a zonal flow with wavelength $k_x \pi G \Sigma_0 / \Omega^2 = 2.0$ was imposed in the initial conditions and was seen to grow exponentially due to the axisymmetric instability observed in Section 6.4.1. The outcome of this purely axisymmetric system was compared with one possessing the same parameters and initial conditions, with the exception of the addition of a white noise IC to provide non-axisymmetry to the flow; the results of this comparison are shown in Figure 6.13. The plot

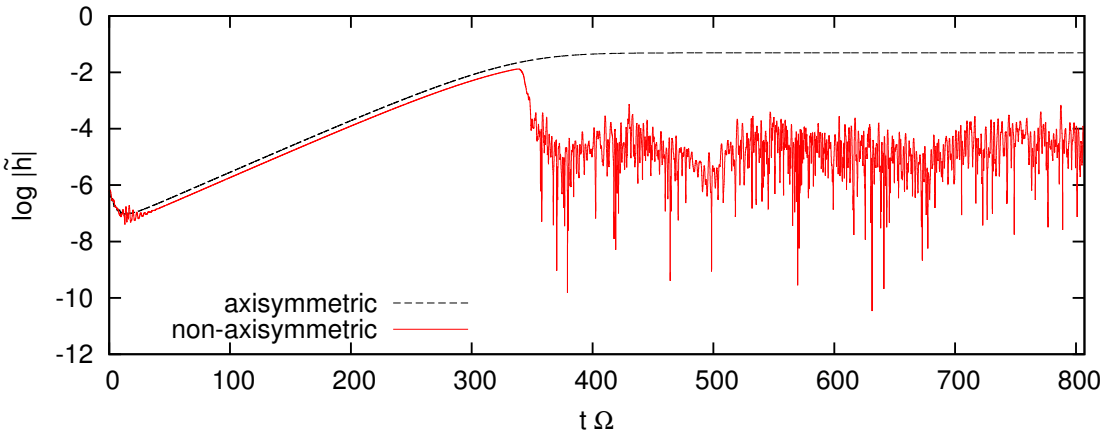


Figure 6.13 – Evolution of the natural log of the initialised zonal flow's amplitude in test runs of axisymmetric (black, dashed line) and non-axisymmetric (red, full) nature with a shear viscosity of $\alpha_s \sim 0.004$ and a Prandtl number of Pr = 3. Following a period of exponential growth where the test runs match, the amplitude of the zonal flow in the non-axisymmetric run is quenched compared to that in the axisymmetric simulation.

shows the evolution of the natural logarithm of the initialised zonal flow's amplitude ($k_x \pi G \Sigma_0 / \Omega^2 = 2.0$) in both test runs. Following a period of exponential growth induced by the slow mode instability, the two runs diverge when entering the non-linear regime, with the non-axisymmetric run (red, full line) showing a quenched amplitude compared to the axisymmetric counterpart (black, dashed), which is nevertheless saturated. This indicates that a non-axisymmetric instability is constraining the amplitude of the zonal flow, similarly to the instability encountered in Chapter 4.

The slow growth of the imposed zonal flows in the test run – despite an initial

shear viscosity matching that in the full runs ($\alpha_s \sim 0.004$) and a larger, more favourable Prandtl number (Pr = 3 compared to an explicit Prandtl number of Pr = 1 in the full runs) – appears to rule out that initial explicit viscosities in the full runs may be causing the emergence and regeneration of zonal flows (e.g. Figure 6.5), which are instead believed to be largely driven by turbulent viscosities.

To confirm the idea of a non-axisymmetric instability disrupting the zonal flows and constraining their properties, and to check whether the non-axisymmetric instability is indeed the same as that encountered in the linear analysis presented in Chapter 4, a kc_s/Ω–A_h stability diagram (like the one in Figure 4.3) was reproduced for the average Toomre parameter value of the test runs ($Q = 1.58$, which is somewhat smaller than that observed in the runs presented in this chapter due to the slight discrepancy in ICs) and $A_e = 0$, with the obtained growth rates being maximised over k_y.

The flow is found to be stable for all A_h values considered for very long zonal flow wavelengths ($2\pi/k \gtrsim 30H$), but shorter wavelength axisymmetric structures are all affected by the instability. It also appears that short enough wavelengths ($2\pi/k \lesssim 2H$) are disrupted by the non-axisymmetric instability for very small amplitudes, which explains the lack of substantial growth in the simulations' axisymmetric structures with wavenumbers larger than $k_x \pi G\Sigma_0/\Omega^2 \gtrsim 2$ (which for $Q \sim 2$ is equivalent to $k_x c_s/\Omega \gtrsim 4$) in Figure 6.5; this is particularly noticeable in the entropy. The plot also features a data point corresponding to the wavenumber considered in the test runs and its average amplitude observed in the non-axisymmetric case from Figure 6.13; this appears to lie very close to the marginal stability contour, therefore confirming the role of the non-axisymmetric instability in constraining the zonal flow amplitudes. Overall, the shape of the growth rate contours appears somewhat similar to those seen in Figure 4.3; the few differences can however be accounted for by the absence of self-gravity in the analysis of Figure 4.3 and the fact that, with a value of $Q^{-1} \simeq 0.65$, Figure 6.14 analyses the growth rates of the high–Q^{-1} gravitational instability from Figure 4.5, rather than the low–Q^{-1} Kelvin-Helmholtz instability, as seen in Figure 6.15.

This presents further analysis on the presence of a non-axisymmetric instability showing – similarly to Figure 4.5 in Chapter 4 – the growth rate contours as a function of the azimuthal wavenumber $k_y c_s/\Omega$ and Q^{-1} for the average zonal flow properties (A_h and kc_s/Ω) from the test run. This shows that the system is completely stable to the Kelvin-Helmholtz instability observed in Figure 4.5 for

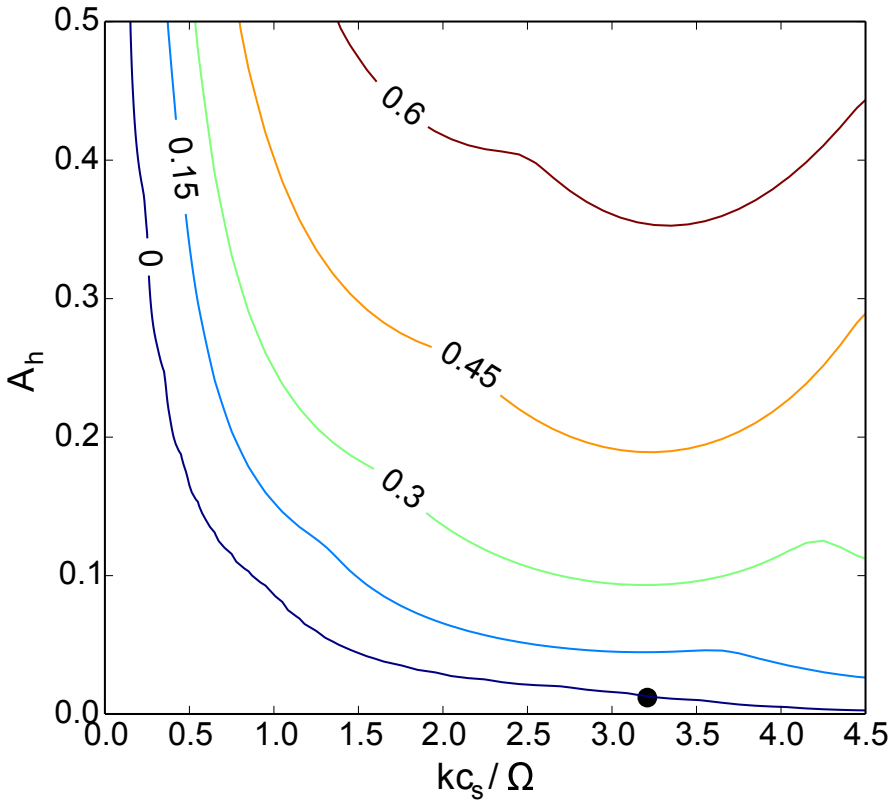


Figure 6.14 – k_y–maximised growth rate contours as a function of the zonal flow properties (amplitude A_h and wavenumber k (in acoustic units)) for $A_e = 0$ and $Q = 1.58$, according to the average Toomre parameter value found in the non-axisymmetric test run. The properties of the zonal flow considered in the test run are plotted as a data point. This lies on the marginal stability contour, hence indicating that the non-axisymmetric instability is acting to constrain the zonal flow amplitude.

low Q^{-1}, but the high–Q^{-1} gravitational instability is active. The black data point corresponding to the test run lies, as expected, relatively close to the marginal stability contour and – taking into consideration that the shape and size of the instability region is substantially affected by changes in A_h and k (the effect of changing the latter being shown in Figure 4.5) – it appears likely the instability is triggered during the test run. Furthermore, the plot shows that a longer-wavelength mode, obtained by considering a more elongated box in the azimuthal direction, might be preferred over the $k_y \pi G \Sigma_0 / \Omega^2 = 0.25$ mode.

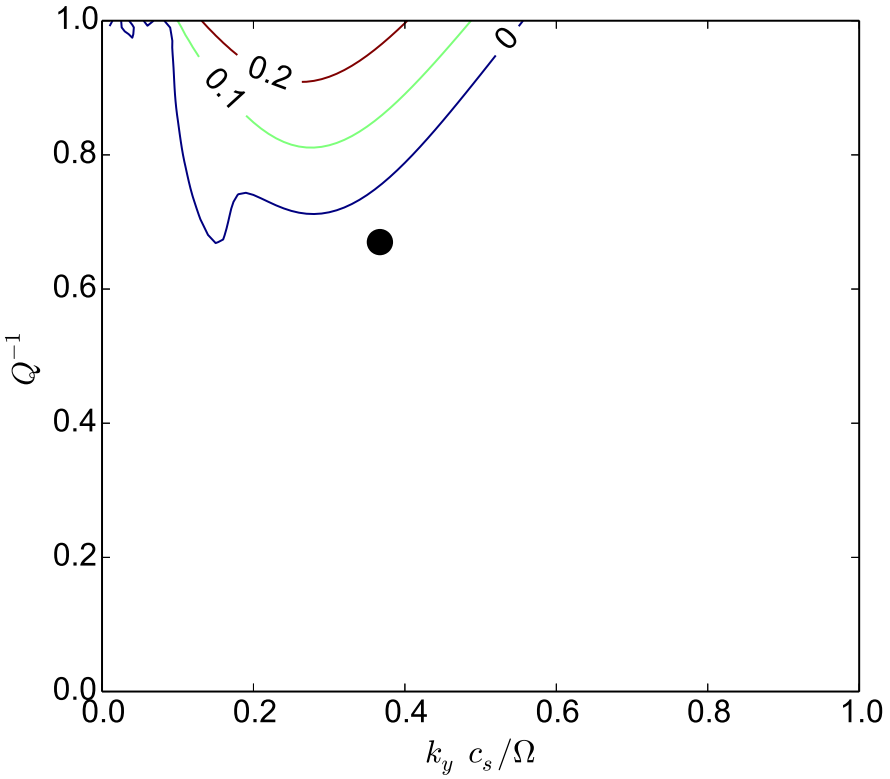


Figure 6.15 – Growth rate contour plot in the $k_y c_s/\Omega$–Q^{-1} plane for the average zonal flow properties from the test run: $kc_s/\Omega \simeq 3.2$ (converting $k\pi G\Sigma_0/\Omega^2 = 2.0$ into acoustic units using the average Q value of the test run, $Q \simeq 1.47$) and $A_h \simeq 0.0105$. The data point represents the average position of the test run's axisymmetric structure in the plane. The shape and size of the unstable region are functions of both k and A_h.

Short-time modulation

While the non-axisymmetric instability acts to constrain the amplitude of the zonal flows, a short-time modulation (on the order of roughly 1 orbit) is also observed in heating events (e.g. Figure 6.2), as well as the longer characteristic timescale of the turbulent regime (roughly $10 - 20$ orbits). This shorter modulation is believed to be caused by the interference between the non-axisymmetric instability and the axisymmetric structure on which it is acting. This implies that the modulation caused by the interaction of the instability on a single zonal flow wavenumber occurs on the shearing wave recurrence timescale (previously used in Section 4)

$$T_{\rm rec} = \frac{k_x}{q\Omega k_y}, \tag{6.16}$$

where k_x is the wavenumber of the axisymmetric structure being considered and k_y the azimuthal wavenumber of the instability.

During periods of cooling, no zonal flow wavelength dominates over the others owing to the flow being stable to the slow mode instability and the previously dominating structures decaying to amplitudes comparable to longer k_x modes, which trigger the non-axisymmetric instability more easily. This causes the period of these modulations to be affected by the interference of the non-axisymmetric instability on any unstable axisymmetric wavenumber present in the disc. This means that the resulting oscillations would not have a regular periodicity (or, in the case where no k_x triggers the non-axisymmetric instability, no modulation would originate from this mechanism). However during heating events usually one or two zonal flow wavenumbers are observed to dominate in the axisymmetric power spectrum, giving these modulations an observable periodicity. An example of this is shown in Figure 6.16, with the modulations in $\overline{Q}$ (**Top**) corresponding to peaks in the dominant structure's PV power spectrum (**Bottom**). As the timescale of the modulations is proportional to the wavenumber of the dominant structure, it is indeed possible to see that their period decreases as the $k_x\pi G\Sigma_0/\Omega^2 = 1.25$ structure becomes prominent from $t\Omega \gtrsim 230$. The timescale observed in $\overline{Q}$ was measured and listed in Table 6.1, with the values being divided into those recorded before and after $t\Omega \sim 230$, as the wavenumber of the dominant structure changes at this point.

For $t\Omega \lesssim 230$, when the $k_x\pi G\Sigma_0/\Omega^2 = 1.50$ structure dominates, a median modulation timescale of $\Delta t^{\rm mod}_{\rm avg}\Omega \simeq 3.919$ is observed. This is $\sim 2\%$ away from the predicted recurrence time of $T_{\rm rec}\Omega = 4.000$ for the given wavenumber. For $t\Omega \gtrsim 230$ the dominant zonal flow's wavenumber is however $k_x\pi G\Sigma_0/\Omega^2 = 1.25$, whose presence should give rise to a modulation with period $T_{\rm rec}\Omega = 3.333$. In this temporal range the runs exhibits a median modulation timescale of $\Delta t^{\rm mod}_{\rm avg}\Omega \simeq 3.340$, which is only 0.2% away from the theoretical value. The accuracy of the observed modulation timescales clearly provides another piece of evidence regarding the presence of a non-axisymmetric instability with $k_y\pi G\Sigma_0/\Omega^2 = 0.25$.

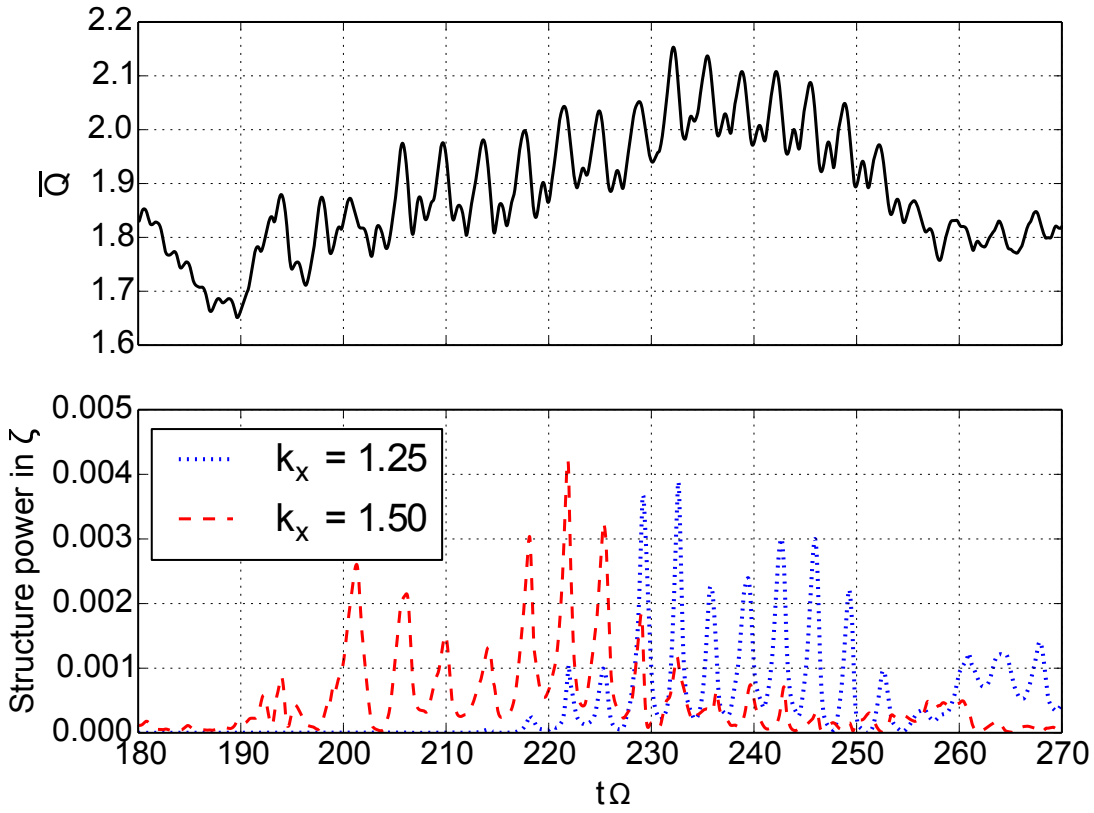


Figure 6.16 – Short-timescale modulation on both $\overline{Q}$ (***Top***) and the PV power spectrum of the two dominant axisymmetric structures' wavenumbers (in units of $\Omega^2/(\pi G\Sigma_0)$, ***Bottom***).

6.4.3 Structure regeneration

After having identified the roles played by the axisymmetric and non-axisymmetric instabilities in the growth and disruption of the axisymmetric structures, the last piece of the puzzle consisted in understanding how these come together. This was only possible by observing the dynamics of the flow in real space, which was carried out thanks to `CASPER`'s ability to output data to `VTK` files. A sequence of such outputs obtained in a run with $\tau_c\Omega = 10$, spanning roughly $t\Omega \simeq 4.5$, is shown in Figure 6.17.

At the beginning of the sequence in (**A**) the sheet presents some degree of axisymmetry on the right side, while on the left the disruption of zonal flows by the non-axisymmetric instability results in the formation of both leading and

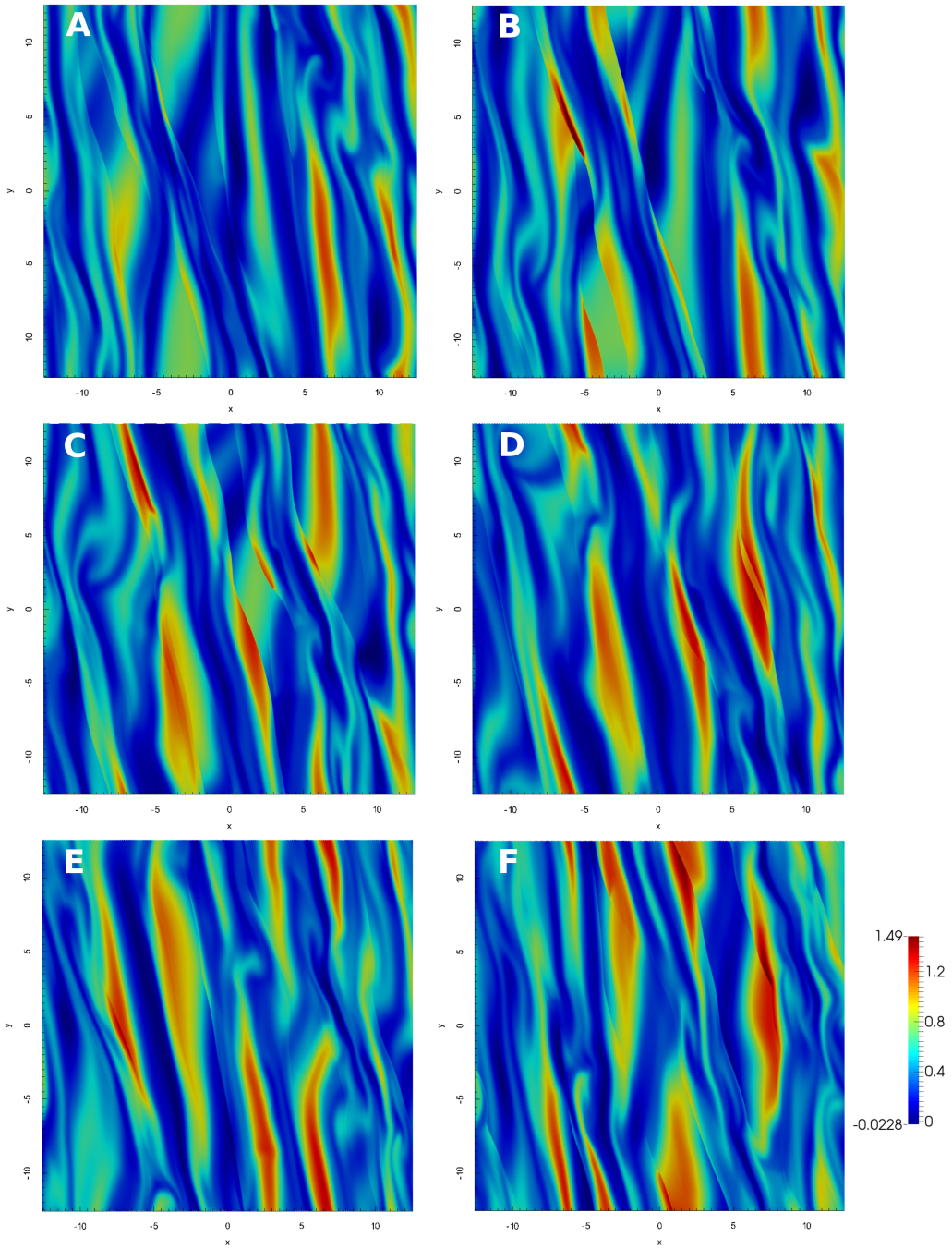


Figure 6.17 – Time sequence showing the regeneration and destruction of the axisymmetric structure in real space in the entropy for $\tau_c = 10\Omega^{-1}$. The time elapsed from (**A**) to (**F**) is $t\Omega \simeq 4.5$.

Table 6.1 – Recorded values for the short-timescale modulation Δt^{mod} in $\overline{Q}$ for both $t\Omega \lesssim 230$ (where the $k_x \pi G \Sigma_0 / \Omega^2 = 1.5$ structure dominates) and $t\Omega \gtrsim 230$ (where the $k_x \pi G \Sigma_0 / \Omega^2 = 1.25$ zonal flow dominates the power spectrum). The median $\Delta t^{\mathrm{mod}}_{\mathrm{avg}}$ of the recorded values is then found and compared to the expected recurrence time T_{rec}.

	$\boldsymbol{t\Omega \lesssim 230}$	$\boldsymbol{t\Omega \gtrsim 230}$
$\boldsymbol{\Delta t^{\mathrm{mod}}\Omega}$	3.933	3.326
	3.934	3.311
	4.029	3.362
	3.855	3.336
	3.415	3.328
	3.904	3.343
		3.365
		3.397
$\boldsymbol{\Delta t^{\mathrm{mod}}_{\mathrm{avg}}\Omega}$	3.919	3.340
$\boldsymbol{T_{\mathrm{rec}}\Omega}$	4.000	3.333

trailing shearing waves. The trailing waves are steepened into shocks (identified by the sharp outlines) in frame (**A**), which travel toward each other and merge at ($x \simeq -6\,\pi G \Sigma_0/\Omega^2$, $y \simeq 5\,\pi G \Sigma_0/\Omega^2$) in frame (**B**), resulting in the production of heat, lighting up the entropy. Due to their curved fronts, the bottom parts of the shocks are still travelling towards each other, and they are observed to merge in frame (**C**), where it is also possible to notice that the axisymmetry on the right has been disrupted and another shock merger is about to occur in the middle of the box. Once this second shock merger takes place in (**D**), the hotspots caused by the mergers are sheared by the flow to form a nearly axisymmetric structure in frame (**E**), which is however once again disrupted by the non-axisymmetric instability in frame (**F**), again generating trailing and leading shearing waves and therefore completing a cycle of regeneration and destruction. This cycle repeats numerous times during the course of a heating event. It is also possible to observe that the amplitude of the zonal flow has grown during the cycle thanks to the action of the slow mode instability.

Figure 6.18 presents a similar analysis by looking at the temporal evolution of some key quantities – such as the axisymmetric and non-axisymmetric power spectra (in entropy and PV, respectively), the Reynolds stress, the kinetic energy of the flow and the Toomre parameter – during a heating event. The slow mode instability causes the zonal flows to grow and, once these have grown sufficiently,

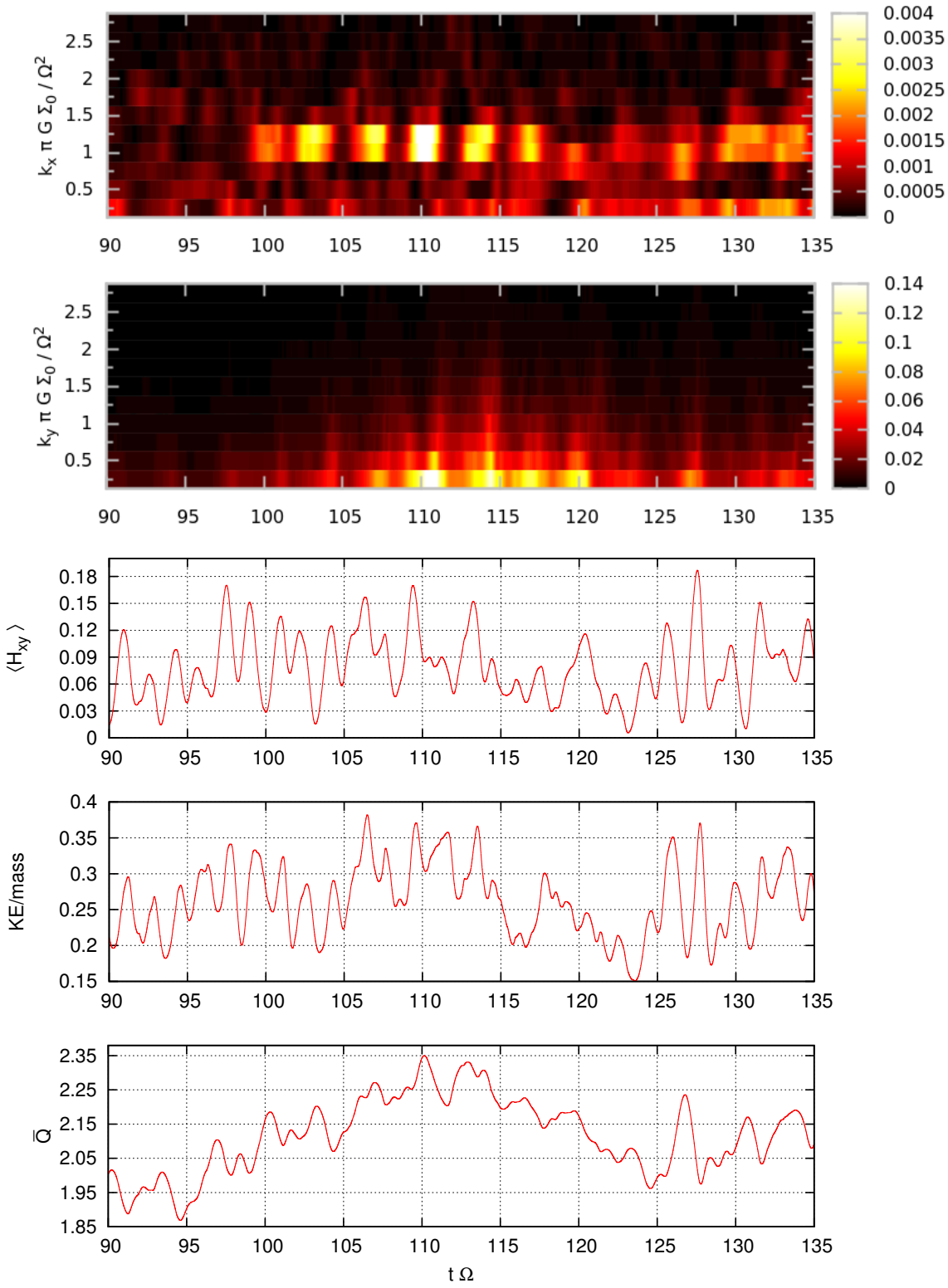


Figure 6.18 – Temporal development of axisymmetric and non-axisymmetric power spectra (in entropy and PV, respectively), Reynolds stress, kinetic energy and Toomre parameter, highlighting the process of gravito-turbulence self-sustenance.

the non-axisymmetric instability is triggered: this disrupts the zonal flow, creating leading and trailing shearing waves. These in turn boost the Reynolds stress, although the peak in $\langle H_{xy} \rangle$ is obtained roughly $1/2 - 1$ orbit after the zonal flow disruption, once the leading wave has undergone its 'swing' to the trailing phase. Energy is therefore removed from the background flow, feeding the turbulent motions; fluid particles then move faster, creating an increase in kinetic energy. This energy is then dissipated into heat as the shocks merge or are damped by viscosity, causing the disc to heat up; this can be observed in $\overline{Q} \propto \sqrt{T}$, where a heating event takes place.

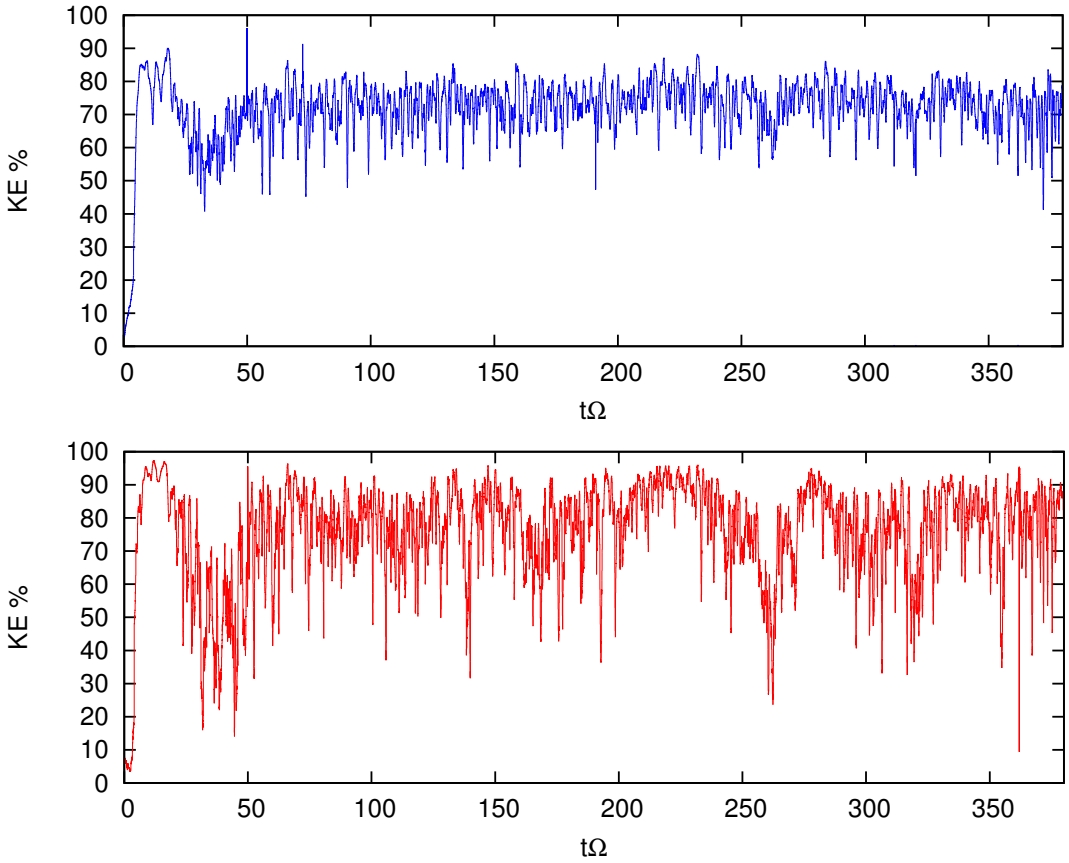


Figure 6.19 – Percentage of the total non-axisymmetric compressive (blue, ***top***) and vortical (red, ***bottom***) trailing kinetic energy present in the $k_y \pi G \Sigma_0 / \Omega^2 = 0.25$ mode.

It is therefore clear that the presence of zonal flows plays a crucial role in the self-sustenance of the gravito-turbulent state and, although the axisymmetric structures are regularly disrupted by the non-axisymmetric instability, the resulting production of leading and trailing shearing waves allows the zonal flows to be

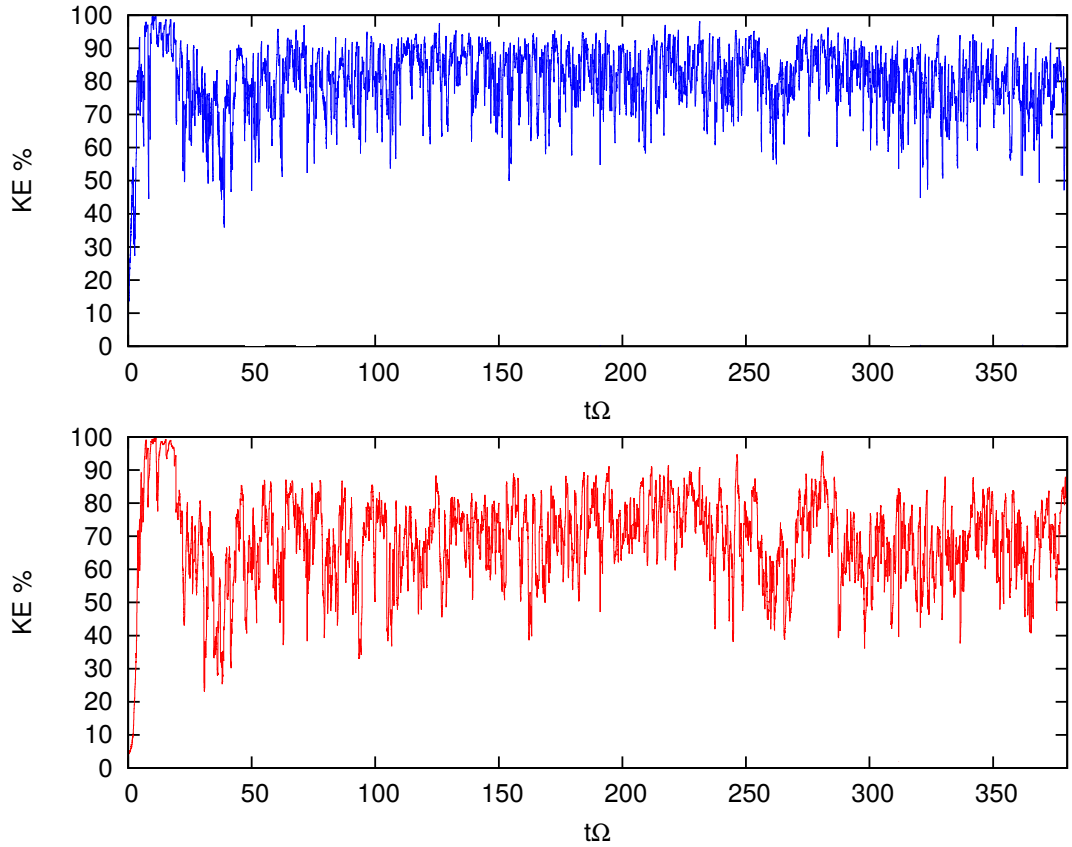


Figure 6.20 – Percentage of the total non-axisymmetric compressive (blue, ***top***) and vortical (red, ***bottom***) leading kinetic energy present in the $k_y \pi G \Sigma_0 / \Omega^2 = 0.25$ mode.

regenerated.

The importance of the non-axisymmetric instability in the gravito-turbulent dynamics is further reinforced by the plots in Figures 6.19 and 6.20, which show the percentage of the total kinetic energy for trailing and leading density waves present in the $k_y \pi G \Sigma_0 / \Omega^2 = 0.25$ mode, respectively. The plots show that this mode, responsible for the non-axisymmetric instability, possesses most of the total shearing wave kinetic energy in both trailing and leading phases, with an average of roughly 75% in each. Both trailing and leading kinetic energies are divided up into compressible and vortical components by calculating the projections of the velocity parallel and perpendicular to the wavenumber vector, as illustrated in Figure 6.21.

The findings of Figures 6.19 and 6.20, in particular the high percentage of total leading kinetic energy provided by the $k_y \pi G \Sigma_0 / \Omega^2 = 0.25$ mode, highlight the importance of the non-axisymmetric instability in the dynamics of the flow. As

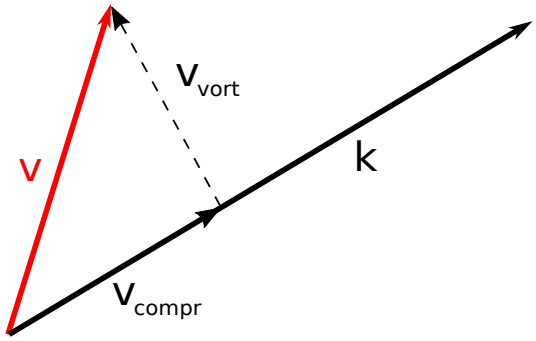


Figure 6.21 – Illustration of the decomposition of the velocity vector into compressible (parallel to $\boldsymbol{k}$) and vortical (perpendicular to $\boldsymbol{k}$) components.

mentioned previously, in order for a gravito-turbulent state to be self-sustaining there must be a regeneration of leading and trailing density waves, which, if coupled to a non-linear feedback, can continuously draw energy from the background flow to sustain gravito-turbulence. It is therefore clear that the non-axisymmetric instability plays a dual and contrasting role in the gravito-turbulent dynamics, as also seen in Figure 6.17. On one hand it disrupts the existing axisymmetric structures when their properties fall within the unstable regions of Figures 6.14 and 6.15, while on the other the zonal flow disruption also provides a source of leading and trailing density waves which directly leads to the regeneration of the zonal flows and the sustenance of gravito-turbulence, as seen in the sequence of Figure 6.17.

6.5 Properties of gravito-turbulence

Just like the stability properties of the system are affected by the choice of some key parameters such as the radiative cooling time and adiabatic index, the properties of the turbulent state can similarly be influenced by some of these choices. While the simulations presented in this chapter were run with very similar initial conditions, some parameters such as cooling time and initial Toomre parameter were varied for a thorough analysis.

Average Q value

An obvious property of the gravito-turbulent state is the average Toomre parameter around which it settles. As seen in Section 6.4.1, the value of $Q_{\rm avg}$ is controlled by the extent of the slow mode instability region, which in turn is affected by

parameters such as the adiabatic index, Prandtl number, shear viscosity and cooling time. While the Prandtl number and the effective shear viscosity vary in a way that cannot be controlled, an interesting analysis is to investigate how the average Toomre parameter depends on the cooling timescale τ_c. As mentioned in Section 6.4.1, and more in detail in Section 5.3.3, increasing the value of τ_c for a fixed value of α and Pr makes the system more unstable to the slow mode instability. However in this instance the value of α_{eff} adapts to different values of τ_c according to $\alpha_{\text{eff}} \propto 1/\tau_c$ (such that all the heating is provided by viscous effects), as the system settles around an average value of α_{eff} which allows it to remain in thermal equilibrium. It is therefore unclear whether changing both α_{eff} and τ_c to maintain thermal balance has an effect on the average value of Q around which the system settles. Such a question is addressed in Figure 6.22, which shows the recorded average Q values (black data points) from runs with various cooling times; the figure also presents the theoretical largest Q (cyan dashed line, obtained from the analysis from Chapter 5) to which the slow mode instability region stretches by means of comparison.

The plot illustrates that varying the value of τ_c has no sizeable impact on the largest Q value unstable to the slow mode instability, as long as all the heat is provided by viscous effects, since the product $\tau_c \alpha_{\text{eff}}$ is kept constant to preserve thermal equilibrium. The results from `CASPER` reflect the theoretical expectation, with the observed average Q showing no clear trend for $\tau_c \leq 13\Omega^{-1}$ and the points being well distributed to either side of the theoretical expectation. The average Q recorded for $\tau_c = 15\Omega^{-1}$ appears however somewhat larger than the rest; this may however be caused by slight inconsistencies between the average values of α_{eff} and Pr_{turb} for this run and the ones used for the slow mode instability. All points remain however within 5% of the theoretical expectation, therefore confirming the role of the axisymmetric instability in the self-sustenance of the gravito-turbulent state.

Characteristic timescale

As partially noticed in Figure 6.2, the self-sustaining state of gravito-turbulence in which the system settles possesses a characteristic timescale Δt on which transient growths succeed one another. Unfortunately in high resolution runs this periodicity is heavily masked by the unpredictable and substantial short-time variations seen in α_{eff} (see Figure 6.7). In lower resolution runs, on the other hand, the ini-

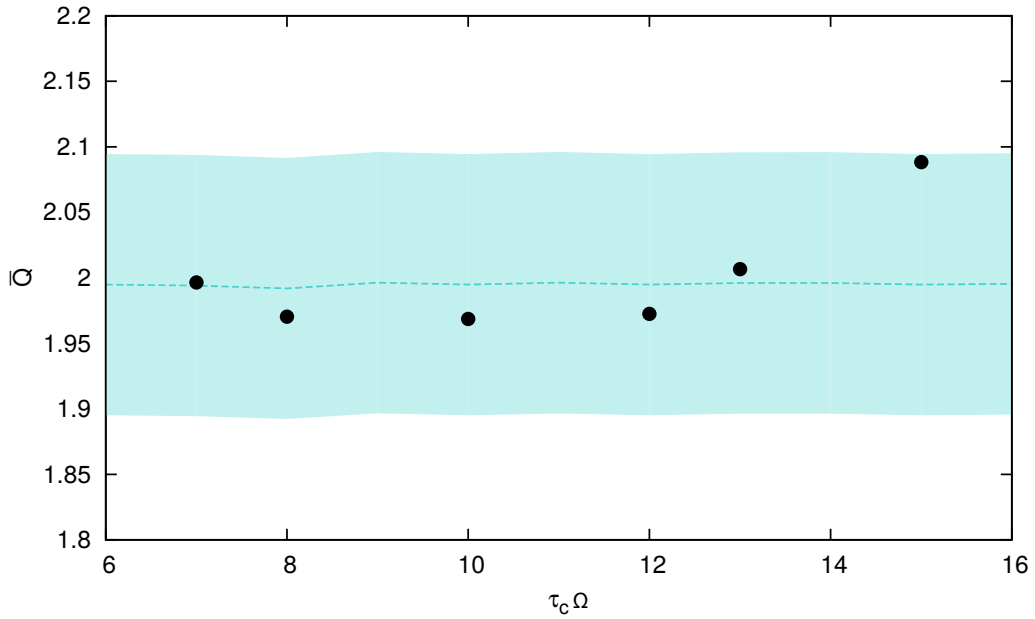


Figure 6.22 – Relationship between the cooling time in the runs and the observed average Toomre parameter in the gravito-turbulent state. The results from CASPER (black data points) are compared with the theoretical maximum value of Q (cyan, dashed line) that the slow mode instability reaches for that cooling time and a value of $\alpha_{\rm eff}$ such that the system is in thermal balance. A fixed Prandtl number $\mathrm{Pr} = 1.9$ is used. Results from runs having the same cooling time have been averaged. The shaded regions represent values within 5% of the theoretical expectation.

tialised viscosity values are too small for the code to maintain numerical stability and CASPER therefore gradually increases them during heating events according to the algorithm illustrated by Equation 3.25. While this does not represent an ideal feature in a numerical simulation, it does mean that the effective viscosity is no longer dominated by the quickly-evolving turbulent viscosity and it has therefore a smoother profile. This allows for a better estimate of the turbulent state's characteristic timescale.

The relationship between the imposed cooling time and the observed turbulent characteristic timescale is investigated in Figure 6.23 for lower resolution 256×256 runs with a box size of $L = 8\pi\,(\pi G\Sigma_0)/\Omega^2$ (red squares) and $L = 10\pi\,(\pi G\Sigma_0)/\Omega^2$ (blue circles). Here the periodicity for a set of parameters is found by calculating the temporal interval between the peaks of any two successive heating events in $\overline{Q}$. The peaks can however extend for a few orbits so their widths are used to estimate the errorbars, which also take into account the regularity of the periodicity.

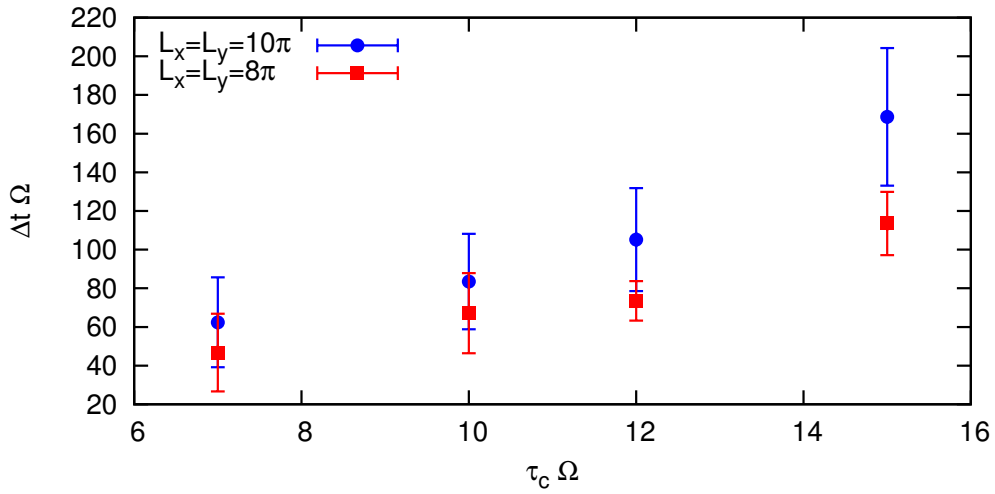


Figure 6.23 – Dependence of the gravito-turbulent characteristic timescale Δt (averaged across the simulation) on the imposed β cooling time τ_c, for $N_x = N_y = 256$, $Q_0 = 1.3$ and with $L = 8\pi\,(\pi G\Sigma_0)/\Omega^2$ (red squares) and $L = 10\pi\,(\pi G\Sigma_0)/\Omega^2$ (blue circles). As expected, a less efficient cooling time results in a longer characteristic timescale, although it is not clear whether Δt increases linearly with τ_c or in a more complex way (e.g. exponentially).

Both sets of points show an increase in the turbulent characteristic timescale with increasing $\tau_c\Omega$, which is easily explained by a disc cooling more slowly taking more time to reach conditions triggering the slow mode instability following a heating event. While the $L = 10\pi\,(\pi G\Sigma_0)/\Omega^2$ runs appear to exhibit a somewhat faster increase of Δt as a function of τ_c than runs with $L = 8\pi\,(\pi G\Sigma_0)/\Omega^2$, the results are comparable and in agreement. It is however difficult to ascertain whether Δt increases linearly with longer cooling times, or whether the relationship is more complex (e.g. exponential).

6.6 Summary

The work presented in this chapter was carried out with the `CASPER` code, which was developed from scratch and whose specifications are set out in Chapter 3, and deals with the ability of a self-gravitating Keplerian flow to settle into a quasi-steady, self-sustaining state by means of a coupling between non-axisymmetric shearing waves and a non-linear feedback. In particular, the mechanism allowing the regeneration of shearing waves, without which the turbulent state cannot be sustained, was at

the centre of this study. For this purpose several high resolution (1024×1024) runs – featuring constant radiative cooling time, viscosities and thermal diffusion – were carried out.

While the runs were initialised with a constant kinematic shear viscosity, this was only equivalent to roughly $\alpha \sim 0.004$ which – due to the absence of an external irradiation term in the internal energy equation – would only provide thermal balance for a radiative cooling timescale of order $\tau_c\Omega \sim 100$. The cooling times used were however an order of magnitude smaller, which meant that the thermally-regulated turbulent state provided an additional source of shear viscosity, which allowed thermal equilibrium. The total observed values of the viscosity, being the sum of the initial and the turbulent values, were in fact found to lie within 10% of the thermal balance theoretical values for a range of cooling times.

Axisymmetric structures in PV and entropy were found to be closely linked with the resulting gravito-turbulent state; these were not applied in the runs' initial conditions, but had developed solely due to the dynamics of the flow. Furthermore, these zonal flows also appeared to be self-sustaining, with their periodic growths accurately coinciding with heating events occurring in the Toomre parameter Q. This appeared to suggest that the zonal flows were playing a role in the sustenance of gravito-turbulence and that an instability, being triggered during these heating events, was responsible for their growth. The slow mode instability (previously discussed in Chapter 5) was indeed found to be responsible for the zonal flow growth, being triggered when the disc reached a low enough temperature and switching off in hotter conditions, despite some assumptions made in the modelling. The instability also determined which zonal flow wavenumbers were active at any one time, with the two wavenumbers centred around the peak of the instability region in the wavenumber–Toomre parameter plane growing fastest.

The zonal flows were however not observed to grow smoothly when unstable to the slow mode instability, instead exhibiting a periodic pattern of oscillations. These were found to be caused by the presence of another instability, this time non-axisymmetric in nature; this instability was found to be the same as the gravitationally-driven one identified in Chapter 4. Its role was identified in constraining the amplitude of zonal flows, with a zonal flow amplitude–wavenumber plot showing how short-wavelength axisymmetric structures are subject to this instability even when possessing a very small amplitude. An analysis of the instability as a function of the azimuthal wavenumber and the Toomre parameter further

showed that were a longer azimuthal wavelength mode to exist due to considering a box more elongated in y, this mode might be preferred. A short-term modulation, observed in the Toomre parameter and other quantities, was also found to be induced by an interference between the non-axisymmetric instability and the zonal flow on which it was acting.

Real space output showed that the state of gravito-turbulence was maintained thanks to the presence of zonal flows and although these were periodically disrupted by the non-axisymmetric instability, this disruption resulted in the generation of leading and trailing shearing waves. These in turn enabled the turbulent state to continuously draw energy from the background flow to maintain itself and their resulting dynamics were seen to regenerate the zonal flow, also thanks to the shearing properties of the flow, on a period of just $4 - 5\Omega^{-1}$. It is important to be aware that the initial viscosities, which are moderately large to ensure the spectral code is stable, could also be playing a role in the self-sustenance of gravito-turbulence.

Lastly, some of the properties of gravito-turbulence were analysed with respect to their initial conditions. While in Chapter 5 the slow mode instability was seen to extend to larger Toomre parameter Q when increasing the cooling time and keeping the shear viscosity fixed, its maximum Q value was seen to be constant when the product of cooling time and viscosity was unchanged to preserve thermal balance. This lack of dependency of Q on the cooling time was also seen in the average Toomre parameters observed in the runs, despite some assumptions in the comparison. Furthermore, the gravito-turbulent regime featured a characteristic timescale which, by means of less noisy lower-resolution runs, was found to increase with increasing cooling time; it was however unclear whether this relationship was of a linear or of a more complex nature.

Chapter 7

Conclusions

7.1 Research Summary

The aim of this thesis was to explore the dynamics and stability of zonal flows in accretion discs, and to establish whether they play a role in the development and sustenance of a gravitationally-driven state of turbulence. Chapters 1 and 2 provided an introductory setting on the types of systems known to host accretion discs and the dynamics of said discs. In particular, emphasis was placed on types of instability that could induce a transport of angular momentum in the outward direction, allowing accretion onto the central object. Furthermore, zonal flows were introduced with evidence of their emergence and their recorded roles in the dynamics of accretion discs.

The linear analysis conducted in Chapter 5 showed that viscous and thermal effects can induce the growth of zonal flows by means of an axisymmetric instability, which is observed to prefer the formation of zonal flows with wavelength comparable to the disc's scale height H. This slow mode instability is particularly effective if the disc is heated mostly by viscous dissipation and if the Prandtl number of the flow exceeds unity. Inefficient cooling was also observed to boost the instability, which was found to stretch to roughly $Q \sim 3$ in the cases considered. The analysis, which was carried out for perturbations of intermediate wavelengths as opposed to the classical theory of viscous and thermal instabilities, resulted in the coupling between entropy and potential vorticity to be of importance in the flow stability. In fact, the coupling was found to aid the slow mode instability, particularly for cases where the Prandtl number exceed unity.

Having established that zonal flows can emerge in viscous discs with cooling and thermal diffusion, a simpler linear analysis (which instead featured no viscous or thermal effects) carried out in Chapter 4 found these zonal flows to be unstable to non-axisymmetric instabilities. In the absence of self-gravity or in weak self-gravitating conditions the zonal flows were found unstable to a Kelvin-Helmholtz instability; this was however stabilised for stronger self-gravity, giving space to a gravitationally-driven instability. The presence of the two instabilities meant that zonal flows were non-axisymmetrically unstable in most of the whole range of the Toomre parameter $0 \leq Q^{-1} \leq 1$. The Kelvin-Helmholtz instability – whose growth rate was found to increase with both increasing zonal flow amplitude and wavenumber, with only large wavelength zonal flows being stable – was also found to partially operate in Rayleigh stable conditions. This means that the KH instability would be the limiting process in constraining the zonal flow's amplitude and wavelength.

The thesis then focused on the results obtained using the pseudo-spectral code `CASPER`, which was developed for the purpose of this work. It represents a fully non-linear, compressible and viscous self-gravitating shear flow, with β-cooling and horizontal thermal diffusion, modelled locally using the shearing sheet approximation. Chapter 3 provided the full set of non-linear equations solved, as well as detailing the specifications and features of the code. Several tests for the soundness of the method were also presented.

In the simulations ran with `CASPER`, which were initialised only with a random white noise perturbation intended to mimic the residual effects of turbulence, the system was seen to settle in a self-sustaining state of gravito-turbulence. Although small constant kinematic viscosities and thermal diffusion were used in the initial conditions to stabilise the flow, their values meant that the system was not in thermal equilibrium. Settling in a self-sustaining state with a roughly constant Q however showed that the system had eventually reached thermal balance thanks to turbulent contributions to the viscous parameters. These were found to be roughly an order of magnitude larger than the initial viscosity values. This gravito-turbulent regime, which oscillated around $Q \approx 2$, was found to be strongly linked to the dynamical evolution of zonal flows. These axisymmetric structures, which emerged spontaneously, were observed to continuously grow and decay as the instabilities described in Chapters 4 and 5 played central roles in the self-sustenance of the turbulent state.

7.2 Implications and observations

The findings from the CASPER runs suggest that zonal flows – usually thought of as products of other mechanisms (e.g. MRI) – could instead play a central role in the development and sustenance of gravito-turbulence, although it is not currently understood whether this requires specific conditions to occur.

The importance of zonal flows dynamics observed in the results of this thesis also implies that the self-sustenance of the gravito-turbulence regime is regulated by the slow mode instability and the disruptive non-axisymmetric instabilities. This may explain to some degree the disagreement found in the literature on the average Q value in the gravito-turbulent state. In fact, although the analysis in Section 6.5 shows that the average Q stays constant while changing the cooling timescale for a system in thermal equilibrium, as long as the fraction of viscously generated heat and Prandtl number are kept fixed, the variation of these last two parameters is likely to impact on the value of Q around which the turbulent state settles.

It appears likely that real accretion discs are affected by the slow mode instability, as this is not driven directly by the values of the shear viscosity and thermal diffusion coefficients, but rather by the Prandtl number and the fraction of heat generated by viscous effects. Real discs' turbulent Prandtl numbers are thought to hold very similar values to the ones used here, with $\mathrm{Pr} \sim 1$.

Zonal flows are known to act like dust traps – particularly effective when dust particles are small enough to be strongly coupled to the gas – so observations in suitable wavelength regimes (e.g. ultraviolet/infrared) could potentially investigate the presence of zonal flows in self-gravitating discs by imaging dust distributions.

7.3 Future developments

The analyses carried out in this thesis present a wide scope of improvement. The non-linear code CASPER can in particular be made even more thorough and comprehensive. This would include considering a more realistic form of cooling (rather than the constant β cooling applied) and viscous parameters which are a function of surface density, radius and/or temperature. Additionally, although the code allows for the introduction of a softening length (as discussed in Section 3.3.5), time constraints made it impossible to use this feature and test the effects of changing of its value.

Although it appears convincing that zonal flows can play a primary role in the self-sustenance of gravito-turbulence, it is not yet clear whether this occurs only for certain sets of conditions or universally. It would therefore be of interest to expand the analysis to include a larger choice of cooling timescales, viscosities, initial Toomre parameter values and initial conditions.

An ulterior addition to the code would be the introduction of a further equation governing the dynamics of dust particles. This would allow a much more thorough analysis of disc dynamics, potentially with further implications due to the presence of zonal flows in gravito-turbulent regimes, and also allow analyses into planet formation or fragmentation to be carried out.

Lastly, although some runs were carried out with different box sizes and resolution – with the results found to be qualitatively unchanged – a more thorough investigation on these parameters would be beneficial.

Appendices

Appendix A

Burgers' equation as a test for the CASPER code

As briefly discussed in Chapter 3, Burgers' equation represented an ideal first computational test for the development of the `CASPER` code thanks to the presence of both diffusive and advective effects.

The full form of the equation is given by

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu_b \frac{\partial^2 u}{\partial x^2}, \tag{A.1}$$

where in this instance the bulk viscosity ν_b was chosen to be a time-independent constant, $\nu_b \frac{\partial^2 u}{\partial x^2}$ is the diffusive term and $u\frac{\partial u}{\partial x}$ is the non-linear advective term. This can be simplified to its inviscid form

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0, \tag{A.2}$$

although the tests carried out aimed at solving the equation's full form.

Figure A.1 shows the temporal evolution of the velocity u profile between $t = 0.001$ (black line) and $t = 30.0$ (magenta), for a fixed timestep $\Delta t = 10^{-3}$ and a bulk viscosity $\nu_b = 0.05$ (with the units such that $u_0 = 1$ and $L = 10$); The solutions given by the viscous Burgers' equation for an initial velocity profile $u(t = 0) = \sin(2\pi x/L)$ and $u(t = 0) = \mathrm{e}^{-[2(2-x)]^2}$ are shown in the top and bottom panels, respectively.

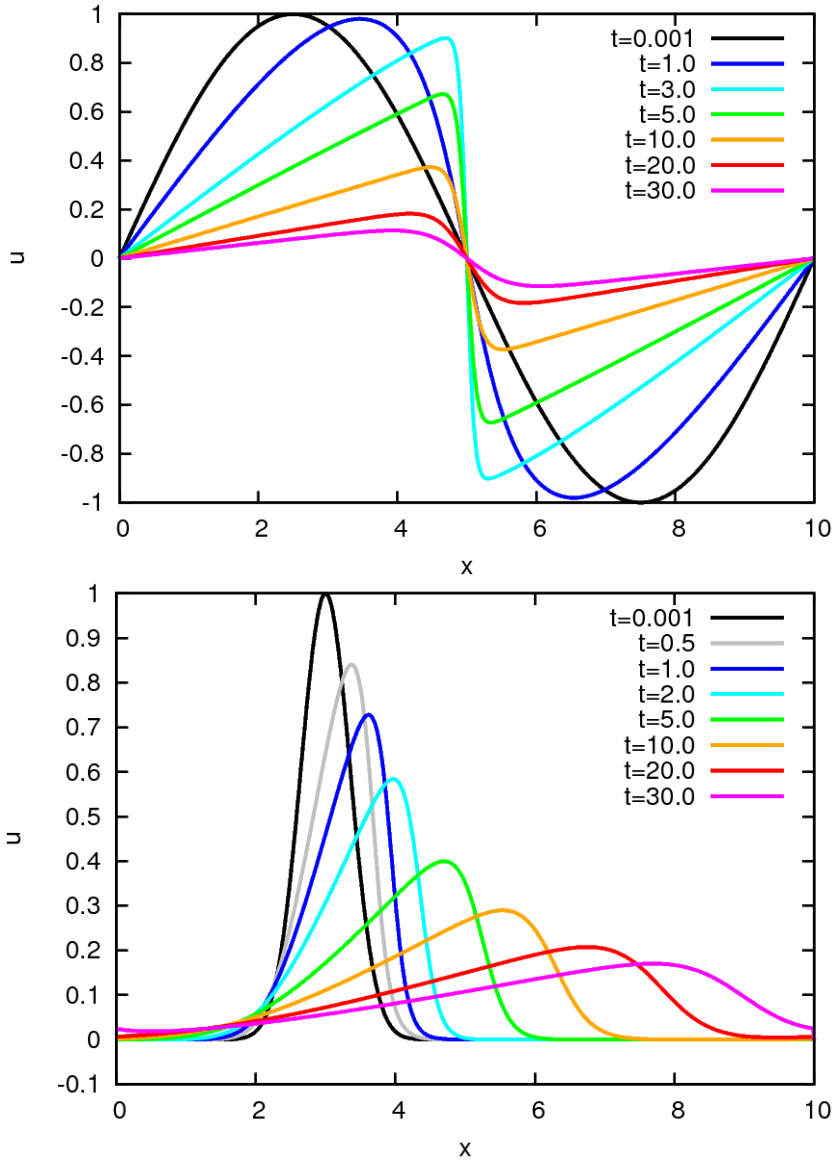


Figure A.1 – Solutions to the viscous Burgers' equation for a range of simulation times with $\nu_b = 0.05$ and $\Delta t = 10^{-3}$ for ***top:*** $u(t = 0) = \sin(2\pi x/L)$ and ***bottom:*** $u(0) = \mathrm{e}^{-[2(2-x)]^2}$.

A.1 Integrating factor method

The first alternative method used to solve Burgers' equation was the integrating method. The use of this method meant that the linear viscous term was dealt with exactly, therefore removing its constraints on the accuracy and stability of the simulation. Taking the Burgers' equation in Fourier space

$$\frac{\partial \widetilde{u}}{\partial t} + \nu_b k^2 \widetilde{u} = -\widetilde{F}, \tag{A.3}$$

where in this case $\widetilde{F}$ represents the Fourier transform of the non-linear term $\frac{\partial u}{\partial x}u$ (which is evaluated in real space so as to bypass the need of a convolution), we introduce the integrating factor I_f, which is defined as

$$I_F = \mathrm{e}^{\nu_b k^2 t}, \tag{A.4}$$

which is multiplied to both sides of Equation A.3 to obtain

$$\mathrm{e}^{\nu_b k^2 t}\frac{\partial \widetilde{u}}{\partial t} + \nu_b k^2 \mathrm{e}^{\nu_b k^2 t}\widetilde{u} = -\mathrm{e}^{\nu_b k^2 t}\widetilde{F}. \tag{A.5}$$

We then introduce the variable $\widetilde{U}$, which is defined as $\widetilde{U} = \widetilde{u}\, I_f = \widetilde{u}\, \mathrm{e}^{\nu_b k^2 t}$ such that its time derivative is

$$\frac{\partial}{\partial t}\left(\widetilde{u}\, \mathrm{e}^{\nu_b k^2 t}\right) = \nu_b k^2 \mathrm{e}^{\nu_b k^2 t}\widetilde{u} + \mathrm{e}^{\nu_b k^2 t}\frac{\partial \widetilde{u}}{\partial t} = \nu_b k^2 \widetilde{U} + \mathrm{e}^{\nu_b k^2 t}\frac{\partial \widetilde{u}}{\partial t}. \tag{A.6}$$

Substituting the expression $\frac{\partial}{\partial t}\left(\widetilde{u}\, \mathrm{e}^{\nu_b k^2 t}\right)$ obtained from Equation A.6 into Equation A.5 gives the inviscid form of Burgers' equation

$$\frac{\partial \widetilde{U}}{\partial t} + I_f \widetilde{F} = 0. \tag{A.7}$$

A.2 Cole-Hopf transformation method

Yet another method that was used to test the accuracy of the solution to Burgers' equation obtained in Section 3.5.1 is the Cole-Hopf transformation; this linearises Burgers' equation, therefore making its solution exact.

In order to make use of the Cole-Hopf method, the velocity u must be transformed according to (Hopf, 1950; Cole, 1950)

$$u = -2\nu_b \frac{1}{\varphi}\frac{\partial \varphi}{\partial x}, \tag{A.8}$$

or alternatively the new function φ can be defined as

$$\varphi(x,t) = \exp\left[-\frac{1}{2\nu_b}\int u\ \mathrm{d}x\right]. \tag{A.9}$$

Applying this transformation to Burgers' equation turns it into the linear equation

$$\frac{\partial \varphi}{\partial t} = \nu_b \frac{\partial^2 \varphi}{\partial x^2}, \tag{A.10}$$

which coincides with the diffusion equation, after redefining φ as $\varphi = \varphi \mathrm{e}^{-\int C \mathrm{d}t}$ so as to remove the constant arising from the integration with respect to x.

Figure A.2 shows the solution to Burgers' equation using the Cole-Hopf transformation. The resulting evolution of the initial conditions is identical to that

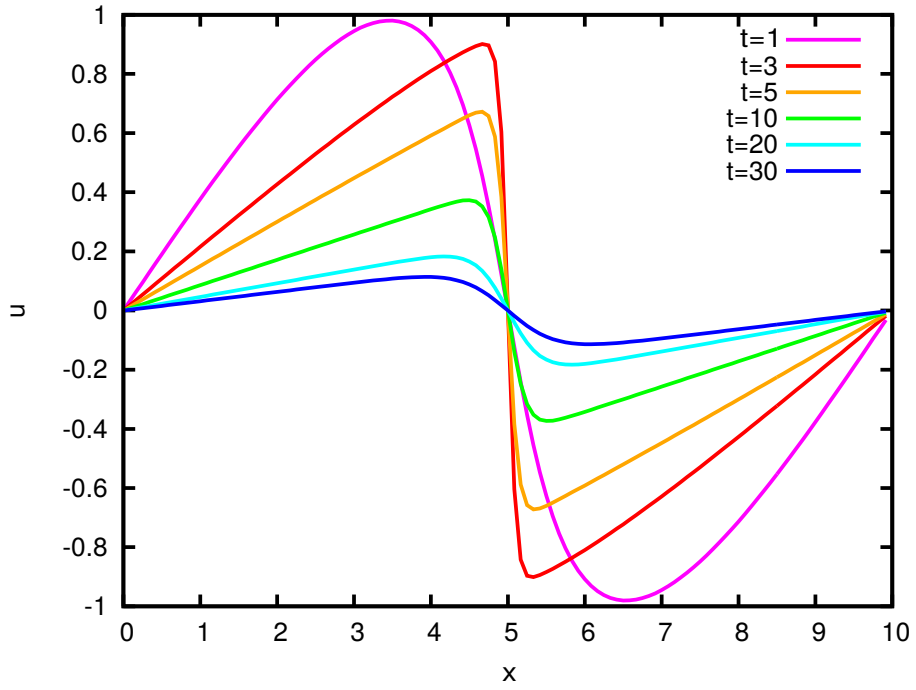


Figure A.2 – Solution to the Burgers' equation using the Cole-Hopf transformation method, for $\nu_b = 0.05$, $\Delta t = 10^{-3}$ and $u(0) = \sin(2\pi x/L)$.

observed in the top panel of Figure A.1, hence corroborating the previous results.

A.3 Riemann invariant method

As the `CASPER` code was expanded from 1D to a 2D isothermal simulation, another test of its output was carried out.

By setting $v' = 0$ and removing any viscosity contribution (with the effects of

shear and the Coriolis force not yet taken into account at this stage of the code development) the following equations for the disturbances in u and h are obtained:

$$\frac{\partial u'}{\partial t} + u' \frac{\partial u'}{\partial x} = -c_{\text{iso}} \frac{\partial h'}{\partial x}, \tag{A.11}$$

$$\frac{\partial h'}{\partial t} + u' \frac{\partial h'}{\partial x} = -c_{\text{iso}} \frac{\partial u'}{\partial x}. \tag{A.12}$$

By adding these two equations together and subtracting the latter from the former, one obtains

$$\frac{\partial}{\partial t}(u' + h') + (u' + c_{\text{iso}}) \frac{\partial}{\partial x}(u' + h') = 0, \tag{A.13}$$

$$\frac{\partial}{\partial t}(u' - h') + (u' - c_{\text{iso}}) \frac{\partial}{\partial x}(u' - h') = 0. \tag{A.14}$$

Each of the two equations represents a wave ($u' + h'$ and $u' - h'$) being propagated at a certain speed ($u' + c_{\text{iso}}$ and $u' - c_{\text{iso}}$, respectively).

The two new quantities $u' + h'$ and $u' - h'$ can be defined as the Riemann invariants $R_{\pm} = u' \pm h'$, so that Equations A.13 and A.14 can be re-written as

$$\frac{\partial R_+}{\partial t} + (u' + c_{\text{iso}}) \frac{\partial R_+}{\partial x} = 0, \tag{A.15}$$

$$\frac{\partial R_-}{\partial t} + (u' - c_{\text{iso}}) \frac{\partial R_-}{\partial x} = 0. \tag{A.16}$$

As a simple wave is being considered, only one of the Riemann invariants can be constant everywhere in the domain; in this case it was chosen to be R_-. R_+, on the other hand, is constant along curves with slope $w = u' + c_{\text{iso}}$ (which are therefore straight lines, as the slope is constant) in the space-time diagram (called characteristic lines, with w being the characteristic velocity).

The two original quantities can then be redefined as

$$u' = w - c_{\text{iso}}, \tag{A.17}$$

$$h' = w - c_{\text{iso}} - R_-, \tag{A.18}$$

meaning that

$$R_+ = u' + h' = 2w - 2c_{iso} - R_-. \tag{A.19}$$

Applying this last expression to Equation A.15 gives the inviscid Burgers' equation

$$\frac{\partial w}{\partial t} + w\frac{\partial w}{\partial x} = 0, \tag{A.20}$$

as R_- was initially taken to be constant.

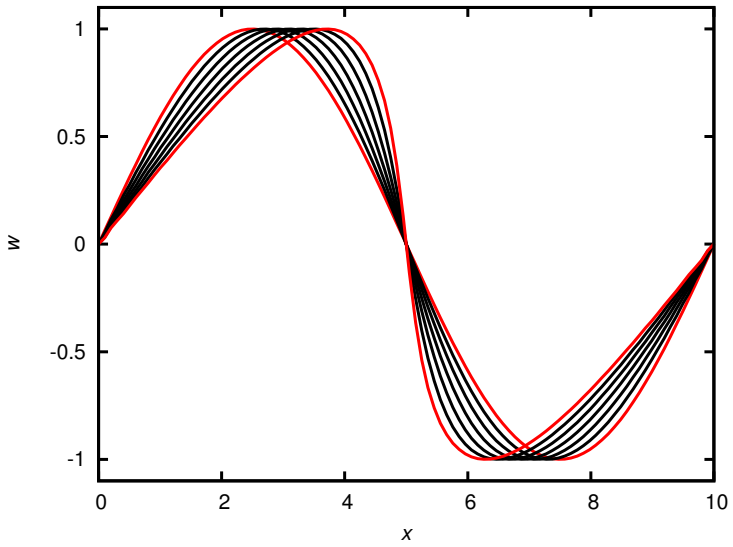


Figure A.3 – Solution to the inviscid Burgers' equation obtained by means of the Riemann invariant method.

Figure A.3 shows the solution to Equation A.20 for the quantity w, with the red curves indicating the solution at the smallest and largest integration times employed. The wave was not integrated for longer times as – due to the absence of viscosity – spurious oscillations due to the Gibb's phenomenon would start to appear on the solution.

Appendix B

Shearing wave ladder truncation

As mentioned in Chapter 4, the number of modes selected when considering the ladder of shearing waves has an impact on how accurately the system at hand is described, with the ladder ideally featuring an infinite amount of modes. While this is clearly impossible to achieve computationally, it is important to understand the implications of choosing a finite amount of modes. As it is shown below, truncating the ladder does not affect the growth rates obtained, therefore validating the analysis presented in Chapter 4.

B.1 Effect on energy conservation

Truncating the ladder of shearing waves means that some energy is lost from the system at every timestep of the analysis, as modes shearing beyond the maximum allowed wavenumber would be dropped. This in turn breaks the time-reversal symmetry of the set-up. In order to explore how the number of modes considered affects the energy conservation of the system, a plot for the growth rate λ/Ω as a function of the azimuthal wavenumber is obtained for various values of N, with the total number of modes given by $2N+1$. The plot, shown in Figure B.1, shows the evolution of the growth rates over $k_y c_s/\Omega$ for $A_h = 0.05$, $kc_s/\Omega = 8$ and $N = 2$ (dotted, magenta line), $N = 4$ (blue, dot-dashed), $N = 6$ (cyan, dashed) and $N = 8$ (red, full); a comparison with the analysis carried out by Lithwick (2007) is also shown (black, thick dashed line). While the positive growth rate values do not seem particularly affected by the different N values (which is further explored in Appendix B.2) and match those found by Lithwick (2007) for $k_y c_s/\Omega \lesssim 1$, the

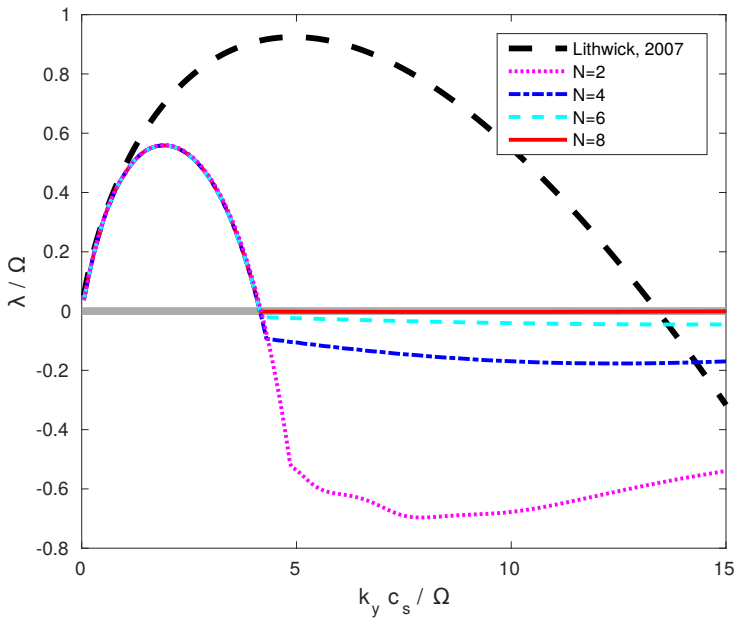


Figure B.1 – Growth rate λ/Ω as a function of the azimuthal wavenumber for the parameter values $kc_s/\Omega = 8$, $A_h = 0.05$ for different amounts of total modes $2N+1$ in the shearing wave ladder. In particular, the cases of $N = 2$ (magenta, dotted line), N = 4 (blue, dot-dashed), N = 6 (cyan, dashed), N = 8 (red, full) are looked at. The growth rates are also compared to those from Lithwick (2007) (thick black, dashed line). A line for $\lambda/\Omega = 0$ is also plotted as a means of reference. It is possible to observe that increasing the total number of modes in the ladder minimises the amount of energy loss from the decaying modes ($\lambda/\Omega \leq 0$). On the other hand the growth rates responsible for the growth of the ladder (i.e. $\lambda/\Omega > 0$) appear already well converged with $N = 2$ and do not seem affected by the increase in the total number of modes.

negative growth rate values associated with decaying modes appear quenched as N is increased, thus reducing the energy lost from the system. The decaying modes' growth rates however appear to converge to $\lambda/\Omega \to 0$ fairly quickly, with $N = 6$ already providing an energy loss that can be considered negligible.

Increasing the size of the ladder from 5 ($N = 2$) to 17 ($N = 8$) modes does nevertheless result in a substantial increase in computational time, which was an issue particularly when self-gravity was introduced in the analysis. Therefore, given that said analysis was based on growing modes – rather than neutral/decaying modes – the value of N used was minimised according to how it affected the growth rates accuracy, as studied below.

B.2 Effect on growth rates

As mentioned earlier, Figure B.1 can also be used to extract important information on the accuracy of the growth rates as a function of N; this is done in Figure B.2 by considering the azimuthal wavenumber range of $0 \leq k_y c_s/\Omega \leq 4$.

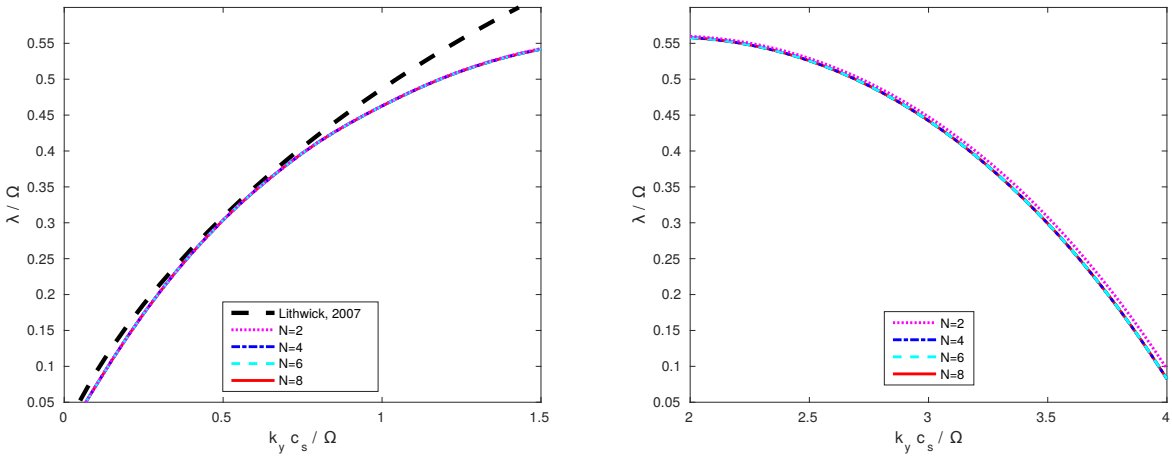


Figure B.2 – Zoomed-in sections of Figure B.1 for the regions $0 \leq k_y c_s/\Omega \leq 1.5$ ***(a)*** and $2 \leq k_y c_s/\Omega \leq 4$ ***(b)***. The growth rates plotted are well converged even for low N values in both k_y ranges; only a small discrepancy can be seen between the $N = 2$ and the others in ***(b)***, but it is of negligible effect on the result.

Unlike the previous analysis for the amount of energy lost from the system due to the truncation of the ladder, the positive growth rates do not seem substantially affected by decreasing the number of modes considered. Indeed, in both Figures B.2a and B.2b, the growth rates appear well converged for all values of N analysed. Indeed the ascending part of the growth rate (Figure B.2a) shows no visible discrepancy between the $N = 2$ case and those with higher N, while in the descending section (Figure B.2b) only a small discrepancy is observed between the $N = 2$ line and the others. Due to the focus of the analysis conducted in Chapter 4 being solely on the growing modes and their growth rates, the outcome of Figure B.2 meant that N could be minimised (often to $N = 2$ or $N = 3$) with the confidence of the result obtained being largely independent of N. This was indeed tested throughout the process of obtaining the results shown in Chapter 4, with the outcomes clearly showing this to be the case.

B.3 Undefined neutral modes

The plot in Figure B.1 shows that, for the chosen axisymmetric structure wavenumber of $kc_s/\Omega = 8$, the amount of energy lost from the system decreases with a higher amount of modes considered within the truncated shearing wave ladder, with the energy loss quickly converging to zero for $N \simeq 6 - 8$. The convergence mentioned above is however not present when an axisymmetric structure of longer wavelength is considered; the $kc_s/\Omega = 2$ instance does not present any convergence to the growth rates of the decaying modes, even when considering very large values of N ($N \sim 100$). The reason for this is believed to be oscillations in decaying modes operating at the smallest resolved scale. It therefore appears that, while regular neutral modes may be present in incompressible cases (e.g. $kc_s/\Omega = 8$), only neutral modes of a spurious nature exist in compressible conditions (e.g. $kc_s/\Omega = 2$).

It is however worth stressing that the analysis presented in Chapter 4 is not based in any way on the non-converging neutral and decaying modes mentioned here. Instead, the analysis is fully based on growing modes, which are found to be fully converged for all values of N and kc_s/Ω examined, including in the compressible conditions mentioned above.

Appendix C

Derivation and verification of Parseval's theorem

Verifying the validity of Parseval's theorem represented a useful test for the code, as the theorem was used to establish the gravitational stress in `CASPER`'s runs. Carrying out this verification is possible thanks to the spectral nature of the code that permits the analysis of quantities in both real and Fourier spaces. The theorem is first derived by considering the following real space quantities

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k}} \tilde{f}(\boldsymbol{k}) \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{C.1}$$

$$g(\boldsymbol{x}) = \sum_{\boldsymbol{k}} \tilde{g}(\boldsymbol{k}) \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{C.2}$$

where both $\tilde{f}(\boldsymbol{k})$ and $\tilde{g}(\boldsymbol{k})$ are Hermitian functions, therefore obeying the property

$$\tilde{f}(-\boldsymbol{k}) = \tilde{f}(\boldsymbol{k})^*, \tag{C.3}$$

with $\tilde{f}(\boldsymbol{k})^*$ being the complex conjugate of $\tilde{f}(\boldsymbol{k})$.

The product of the two real space quantities is therefore given by

$$\begin{aligned}
\int_0^{L_x}\int_0^{L_y} f(x)g(x)\,\mathrm{d}x\,\mathrm{d}y &= \int_0^{L_x}\int_0^{L_y} \sum_{\boldsymbol{k}}\sum_{\boldsymbol{k}} \tilde{f}(\boldsymbol{k})\,\tilde{g}(\boldsymbol{k}')\mathrm{e}^{\mathrm{i}(\boldsymbol{k}+\boldsymbol{k}')\cdot\boldsymbol{x}}\,\mathrm{d}x\,\mathrm{d}y \\
&= \sum_{\boldsymbol{k}} \tilde{f}(\boldsymbol{k})\,\tilde{g}(-\boldsymbol{k})\,L_xL_y \\
&= \sum_{\boldsymbol{k}} \tilde{f}(\boldsymbol{k})\,\tilde{g}(\boldsymbol{k})^*\,L_xL_y,
\end{aligned} \tag{C.4}$$

which represents Parseval's theorem for the given quantities.

The theorem is easily applicable to the 2-dimensional pressure $P = (\gamma - 1)\Sigma e$, for which Parseval's theorem states

$$\frac{(\gamma - 1)}{L_x L_y} \int_0^{L_x} \int_0^{L_y} \Sigma(x) e(x) \, \mathrm{d}x \, \mathrm{d}y = (\gamma - 1) \sum_{\boldsymbol{k}} \tilde{\Sigma}(\boldsymbol{k}) \, \tilde{e}(\boldsymbol{k})^*. \tag{C.5}$$

It is important to notice that while the properties of the `FFTW` method used allowed the use of only half the total $\boldsymbol{k}$ points available (i.e. $k_y \geq 0$), the sum in Equation C.5 requires the consideration of every point in Fourier space. This can however be readily rectified by means of the Hermitian function property defined in Equation C.3; this yields

$$\sum_{\boldsymbol{k}} \tilde{\Sigma}(\boldsymbol{k}) \, \tilde{e}(\boldsymbol{k})^* = \sum_{k_x} \sum_{k_y \geq 0} \tilde{\Sigma}(\boldsymbol{k}) \, \tilde{e}(\boldsymbol{k})^* + \sum_{k_x} \sum_{k_y < 0} \tilde{\Sigma}(\boldsymbol{k})^* \, \tilde{e}(\boldsymbol{k}). \tag{C.6}$$

The validity of Parseval's theorem is tested and verified in Figure C.1, which shows the 2D pressure in both real (top) and Fourier (bottom) space. The Fourier space pressure is – as predicted by the theorem – equal to the real space quantity when the latter is normalised by the area of the shearing sheet, as outlined by Equation C.5.

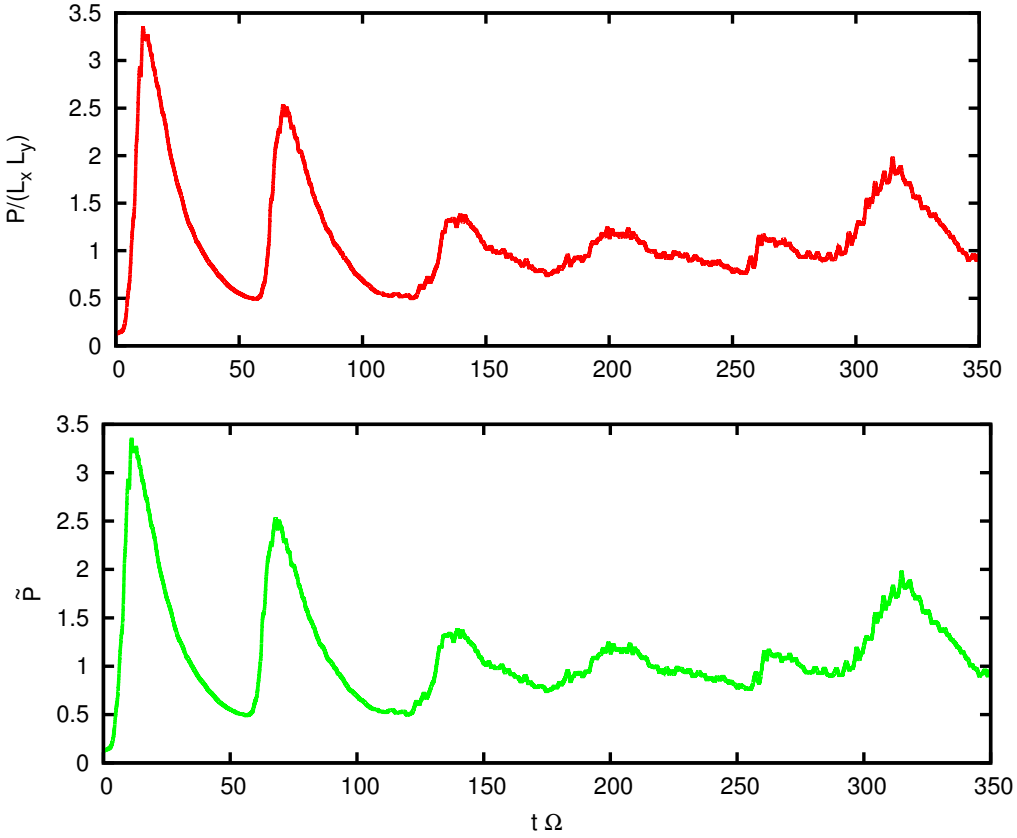


Figure C.1 – Verification of Parseval's theorem by calculating the total pressure of the flow in both real space and Fourier space. The total pressure in $\boldsymbol{k}$-space is equal to the real space pressure divided by the area of the box, as indicated by Parseval's theorem.

Bibliography

Afshordi, N., B. Mukhopadhyay, and R. Narayan (2005). *ApJ* 629:373–382.

Alexander, R.D., C.J. Clarke, and J.E. Pringle (2006). *MNRAS* 369:216–228.

Andrews, S.M., K.A. Rosenfeld, A.L. Kraus, and D.J. Wilner (2013). *ApJ* 771:129.

Armitage, P.J. (1998). *ApJ* 501:L189–L192.

Armitage, P.J. (2011). *Annu. Rev. Astron. Astrophys.* 49:195–236.

Armitage, P.J., M. Livio, and J.E. Pringle (2001). *Mon. Not. R. Astron. Soc.* 324:705–711.

Atkinson, D.H., J.B. Pollack, and A. Seiff (1996). *Science* 272:842.

Baggett, J.S., T.A. Driscoll, and L.N. Trefethen (1995). *Physics of Fluids* 7:883.

Bai, X.-N. and J.M. Stone (2014). *ApJ* 796:31.

Balbus, S.A. (2003). *ARA&A* 41:555.

Balbus, S.A. and J.F. Hawley (1991). *ApJ* 376:214.

Balbus, S.A. and J.C.B. Papaloizou (1999). *ApJ* 521:650–658.

Balbus, S.A., J.F. Hawley, and J.M. Stone (1996). *ApJ* 467:76.

Balmforth, N.J. and D.G. Korycansky (2001). *MNRAS* 326:833–851.

Barranco, J.A. and P.S. Marcus (2005). *ApJ* 623:1157.

Barranco, J.A. and P.S. Marcus (2006). *Journal of Computational Physics* 219:21–46.

Basu, S. (1998). *ApJ* 509:229–237.

Bell, K.R. and D.N.C. Lin (1994). *ApJ* 427:987.

Binney, J. and S. Tremaine (1987). *Galactic Dynamics*. Princeton University Press, Princeton, NJ.

Blumenthal, G.R., L.T. Yang, and D.N.C. Lin (1984). *ApJ* 287:774.

Bode, M.F. (2011). "Classical and Recurrent Nova Outbursts".

Bodo, G., A. Mignone, F. Cattaneo, P. Rossi, and A. Ferrari (2008). *A&A* 487:1.

Boffin, H.M.J., S.J. Watkins, A.S. Bhattal, N. Francis, and A.P. Whitworth (1998). *MNRAS* 300:1189–1204.

Boley, A.C. (2009). *ApJ* 695:L53.

Boley, A.C., A.C. Mejía, R.H. Durisen, K. Cai, M.K. Pickett, and P. D'Alessio (2006). *ApJ* 651:517.

Boley, A.C., T. Hayfield, L. Mayer, and R.H. Durisen (2010). *Icarus* 207:509.

Bonnell, I. and P. Bastien (1992). *ApJ* 401:L31.

Bonnell, I.A., P. Clark, and M.R. Bate (2008). *MNRAS* 389:1556–1562.

Borderies, N., P. Goldreich, and S. Tremaine (1985). *Icarus* 63:406.

Boss, A.P. (1997). *Science* 276:1836–1839.

Boss, A.P. (1998). *ApJ* 503:923–937.

Boss, A.P. (2004). *ApJ* 610:456.

Brandenburg, A., A. Nordlund, R.F. Stein, and U. Torkelsson (1995). *ApJ* 446:741.

Buat-Ménard, V., J.-M. Hameury, and J.-P. Lasota (2001). *A&A*.

Bąkowska, K. and A. Olech (2014). *Acta Astronomica*.

Cai, K., R.H. Durisen, S. Michael, A.C. Boley, A.C. Mejía, M.K. Pickett, and P. D'Alessio (2006). *ApJ* 636:L149.

Cai, T., K.L. Chan, and L. Deng (2011). *Journal of Computational Physics* 230:8698–8712.

Cameron, A.G.W. (1978). *Moon and Planets* 18:5.

Canuto, C., M.Y. Hussaini, A. Quarteroni, and T.A. Zang (2006). *Spectral methods - Fundamentals in Single Domains*. Springer.

Chagelishvili, G.D., J. Zahn, A.G. Tevzadze, and J.G. Lominadze (2003). *A&A* 402:401.

Chan, C.-K., D. Psaltis, and F. Özel (2006). *ApJ* 645:506–518.

Chandrasekhar, S. (1960). *Proceedings of the National Academy of Science* 46:253.

Christensen, U.R. (2001). *Geophysical Research Letters* 28:2553–2556.

Clarke, C.J. (2009). *MNRAS* 396:1066.

Coe, M.J. (2005). *MNRAS* 358:1379–1382.

Cole, J.D. (1950). *Q. Appl. Math.* 9:225–236.

Collins-Souffrin, S. and A.M. Dumont (1990). *A&A* 229:292.

Colwell, J.E., L.W. Esposito, and M. Sremčević (2006). *Geophys. Res. Lett* 33:L07201.

Cossins, P., G. Lodato, and C.J. Clarke (2009). *MNRAS* 393:1157.

Cossins, P., G. Lodato, and C.J. Clarke (2010). *MNRAS* 401:2587.

Costa, B. (2004). *CUBO* 6:1–32.

Dauchot, O. and F. Daviaud (1994). *Phys. Fluids* 7:335.

De Val-Borro, M., P. Artymowicz, G. D'Angelo, and A. Peplinksi (2007). *A&A* 471:1043–1055.

Diamond, P.H., S.I. Itoh, K. Itoh, and T.S. Hahm (2005). *Plasma Phys. Control. Fusion* 47:35.

Done, C., M. Gierliński, and A. Kubota (2007). *Astron. Astrophys. Rev.* 15:1.

Dubrulle, B., O. Dauchot, F. Daviaud, P.-Y. Longaretti, D. Richard, and J.-P. Zahn (2005). *Phys. Fluids* 17:095103.

Dullemond, C.P. and C. Dominik (2005). *Astron. Astrophys.* 434:971–986.

Durisen, R.H., A.P. Boss, L. Mayer, A.F. Nelson, T. Quinn, and W.K.M. Rice (2007). *Protostars and Planets V*:607–662.

Ekaterinaris, J.A. (2005). *Progress in Aerospace Sciences* 41:192–300.

Elmegreen, D.M. (1980). *ApJ* 242:528.

Esner, J.A. (2012). *ApJ* 755:23.

Ferrarese, L. and D. Merritt (2000). *ApJ* 539:L9–L12.

Flannery, B.P. (1974).

Forgan, D. and W.K.M. Rice (2010). *MNRAS* 402:1349.

Forgan, D., W.K.M. Rice, D. Stametellos, and A.P. Whitworth (2009). *MNRAS* 394:882.

Frank, J., A. King, and D. Raine (2002). *Accretion Power in Astrophysics*. Cambridge University Press.

Fricke, K. (1968). *ZAp* 68:317.

Fromang, S. and R.P. Nelson (2005). *MNRAS* 364:L81.

Fromang, S. and J.C.B. Papaloizou (2007). *A&A* 476:1113.

Gallagher, J.S. and S. Starrfield (1978). *Ann. Rev. Astron. Astrophys.* 16:171–214.

Gammie, C. (1999). "Astrophysical Discs". *Astrophysical Discs*. Ed. by J.A. Sellwood and J. Goodman. Astonomical Society of the Pacific.

Gammie, C.F. (1996). *ApJ* 462:725–731.

Gammie, C.F. (2001). *ApJ* 553:174–183.

Gebhardt, K., R. Bender, G. Bower, A. Dressler, S.M. Faber, A.V. Filippenko, R. Green, C. Grillmair, L.C. Ho, J. Kormendy, T.R. Lauer, J. Magorrian, J. Pinkney, D. Richstone, and S. Tremaine (2000). *ApJ* 539:L13.

Gebhardt, T. and S. Grossmann (1994). *Phys. Rev. E* 50:3705.

Gierasch, P.J., A.P. Ingersoll, and D. Pollard (1979). *Icarus* 40:205–212.

Gierliński, M. and C. Done. (2004). *MNRAS* 347:885.

Gill, A.E. (1965). *J. Fluid Mech.* 21:503.

Godon, P. and M. Livio (2000). *ApJ* 537:396–404.

Goldreich, P. and D. Lynden-Bell (1965a). *MNRAS* 130:125.

Goldreich, P. and D. Lynden-Bell (1965b). *MNRAS* 130:97.

Goldreich, P. and G. Schubert (1967). *ApJ* 150:571–588.

Goldreich, P. and S. Tremaine (1978). *ApJ* 222:850.

Goodman, J. and R.R. Rafikov (2001). *ApJ* 552:793–802.

Gorti, U., C.P. Dullemond, and D. Hollenbach (2009). *ApJ* 705:1237–51.

Gottlieb, D. and J.S. Hesthaven (2000). *Journal of Computational and Applied Mathematics* 128:83–131.

Greenhill, L.J. and C.R. Gwinn (1997). *Ap&SS* 248:261.

Grossmann, S. (2000). *Rev. Mod. Phys.* 72:603.

Guillot, T. (1999). *Planet. Space Sci.* 47:1183–1200.

Haghighipour, N. and A.P. Boss (2003). *ApJ* 583:996.

Hardcastle, M.J., D.A. Evans, and J.H. Croston (2007). *Mon. Not. R. Astron. Soc.* 376:1849–1856.

Haring, N. and H.-W. Rix (2004). *ApJ* 604:L89.

Hartmann, L. and S.J. Kenyon (1996). *Annu. Rev. Astron. Astrophys.* 34:207–240.

Hawley, J.F. (2001). *ApJ* 554:534–547.

Hawley, J.F., S.A. Balbus, and W.F. Winters (1999). *ApJ* 518:394.

Hedman, M.M., P.D. Nicholson, H. Salo, B.D. Wallis, B.J. Buratti, K.H. Baines, R.H. Brown, and R.N. Clark (2007). *The Astronomical Journal* 133:2624–2629.

Heimpel, M. and J. Aurnou (2007). *Icarus* 187:540.

Heinemann, T. and J.C.B. Papaloizou (2009a). *MNRAS* 397:52–63.

Heinemann, T. and J.C.B. Papaloizou (2009b). *MNRAS* 397:64–74.

Hellier, C. (2001). *Cataclysmic Variable Stars - How and Why They Vary*. Springer Science & Business Media.

Hewitt, E. and R.E. Hewitt (1979). *Arch. Hist. Exact Sci.* 21(2):129–160.

Hinkley, S., B.R. Oppenheimer, D. Brenner, N. Zimmerman, L.C. Roberts Jr., I.R. Parry, R. Soummer, A. Sivaramakrishnan, M. Simon, M.D. Perrin, D.L. King, J.P. Lloyd, A. Bouchez, J.E. Roberts, R. Dekany, C. Beichman, L. Hillenbrand, R. Burruss, M. Shao, and G. Vasisht (2010). *ApJ* 712:421–428.

Hopf, E. (1950). *Commun. Pure Appl. Math.* 3:201–230.

Johansen, A., A. Youdin, and H. Klahr (2009). *ApJ* 697:1269–1289.

Johnson, B.M. and C.F. Gammie (2003). *ApJ* 597:131–141.

Johnson, B.M. and C.F. Gammie (2005). *ApJ* 635:149.

José, J. and M. Hernanz (1998). *ApJ* 494:680–690.

José, J., J. Casanova, E. Garcia-Berro, M. Hernanz, S.N. Shore, and A.C. Calder (2012). "Classical and Recurrent Nova Models".

Jou, W.-H. and A.C. Mueller (1985). *Flow Technical Report* 325.

Kahabka, P., H.W. Hartmann, A.N. Parmar, and I. Negueruela (1999). *Astron. Astrophys.* 347:L43–L46.

Kant, I. (1755). *Allgemeine Naturgeschichte und Theorie des Himmels*. Ed. by Petersen, Königsberg, and Leipzig.

Käpylä, P.J., A. Brandenburg, M.J. Korpi, J.E. Snellman, and R. Narayan (2010). *ApJ* 719:67–76.

Kato, M.T., K. Nakamura, R. Tandokoro, M. Fujimoto, and S. Ida (2009). *ApJ* 691:1697.

Kato, S. (1978). *MNRAS* 185:629.

Kato, S. and J. Fukue (1980). *PASJ* 32:337.

Kato, S., F. Honma, and R. Matsumoto (1988). *MNRAS* 231:37.

Kelvin, Lord (1887). *Phil. Mag.* 24.

Kenyon, S.J., L.W. Hartmann, Stromm K.M., and S.E. Strom (1990). *Astron. J.* 99:869.

King, A.R. and J.E. Pringle (2007). *MNRAS* 377:L25.

Kitamura, Y., M. Momose, S. Yokogawa, R. tamura, and S. Ida (2002). *ApJ* 581:357.

Klahr, H. and D.N.C. Lin (2001). *ApJ* 554:1095.

Klahr, H.H. and P. Bodenheimer (2003). *ApJ* 582:869.

Klamer, I.J., R.D. Ekers, E.M. Sadler, and R.W. Hunstead (2004). *ApJL* 612:L97.

Kley, W., J.C.B. Papaloizou, and D.N.C. Lin (1993). *ApJ* 416:679.

Kondaxakis, D. and S. Tsangaris (2008). *Fluid Dynamics Research* 40:311–342.

Kondratko, P.T., L.J. Greenhill, and J.M. Moran (2006). *ApJ* 652:136–145.

Kratter, K.M. and G. Lodato (2016). *Annual Reviews* 54.

Kreidberg, L.R., M.A. Wood, and M. Wetzstein (2010). *JSARA* 4:38–41.

Kunz, M.W. and G. Lesur (2013). *MNRAS* 434:2295–2312.

Lamb, F.K., C.J. Pethick, and D. Pines (1973). *ApJ* 184:271–289.

Larson, R.B. (1989). *The Formation and Evolution of Planetary Systems*. Ed. by H.A. Weaver and L. Danly. Cambridge University Press, 31.

Latter, H.N. and G.I. Ogilvie (2006). *MNRAS* 372:1829–1839.

Laughlin, G. and P. Bodenheimer (1994). *ApJ* 436:335.

Laughlin, G. and M. Rozyczka (1996). *ApJ* 456:279.
Lee, W.-K. and F.H. Shu (2013). *ApJ* 756:45.
Lerner, J. and E. Knobloch (1988). *J. Fluid Mech.* 189:117.
Lesur, G. and P.-Y. Longaretti (2005). *A&A* 444:25.
Lesur, G. and G.I. Ogilvie (2010). *MNRAS* 404:L64–L68.
Lesur, G. and J.C.B. Papaloizou (2010). *A&A* 513:60.
Lewin, W.H.G., J. van Paradijs, and E.P.J. van den Heuvel (1997). *X-ray Binaries*. Cambridge University Press.
Li, H., J.M. Finn, R.V.E. Lovelace, and S.A. Colgate (2000). *ApJ* 533:1023.
Lightman, A.P. and D.M. Eardley (1974). *ApJ* 187:L1–L3.
Limaye, S.S. (1986). *Icarus* 65:335–352.
Lin, C.C. and F.H. Shu (1964). *ApJ* 140:646.
Lin, D.N.C. and J.C.B. Papaloizou (1980). *MNRAS* 191:37.
Lin, D.N.C. and J.C.B. Papaloizou (1985). *Protostars and Planets II*. Ed. by D.C. Black and M.S. Matthews, pp. 981–1072.
Lin, D.N.C. and J.C.B. Papaloizou (1996). *Annu. Rev. Astron. Astrophys.* 36.
Lin, D.N.C., G. Laughlin, P. Bodenheimer, and M. Różyczka (1998). *Science* 281:2025–2027.
Lin, M.-K. and J.C.B. Papaloizou (2011a). *MNRAS* 415:1426–1444.
Lin, M.-K. and J.C.B. Papaloizou (2011b). *MNRAS* 415:1445–1468.
Lithwick, Y. (2007). *ApJ* 670:789–804.
Lithwick, Y. (2009). *ApJ* 693:85–96.
Livio, M. and G. Shaviv (1977). *A&A* 55:95–101.
Livio, M. and F. Verbunt (1988). *Mon. Not. R. astr. Soc.* 332.
Lodato, G. (2007). *Nuovo Cimento* 30:293.
Lodato, G. and G. Bertin (2003). *A&A* 398:517–524.
Lodato, G. and W.K.M. Rice (2004). *MNRAS* 351:630.
Lodato, G. and W.K.M. Rice (2005). *MNRAS* 358:1489.
Lodato, G., E. Delgado-Donate, and C.J. Clarke (2005). *MNRAS* 364:L91.
Longaretti, P.-Y. (2002). *ApJ* 576:587–598.
Longaretti, P.-Y. and G. Lesur (2010). *A&A* 516:A51.
Lorén-Aguilar, P. and M.R. Bate (2015). *MNRAS* 453:L78.
Lovelace, R.V.E. and R.G. Hohlfeld (1978). *ApJ* 221:51.
Lovelace, R.V.E. and R.G. Hohlfeld (2013). *MNRAS* 429:529–533.
Lovelace, R.V.E., H. Li, S.A. Colgate, and A.F. Nelson (1999). *ApJ* 513:805.

Lubow, S.H. and G. D'Angelo (2006). *ApJ* 641:526–533.

Lubow, S.H. and S. Ida (2010). *Exoplanets*. Ed. by S. Seager. Tucson: Univ. Ariz. Press, 347.

Lüst, R. (1952). *Z. Naturforsch* 7a:87–98.

Lynden-Bell, D. (1969). *Nature* 223:690–694.

Lynden-Bell, D. and A.J. Kalnajs (1972). *MNRAS* 157:1.

Lynden-Bell, D. and J.E. Pringle (1974). *MNRAS* 168:603.

Lynds, B.T. (1970). "The Distribution of Dark Nebulae in Late-Type Spirals". *The Spiral Structure of Our Galaxy*. Ed. by W. Becker and G. Contopoulos. Springer Netherlands, pp. 26–34.

Lyra, W. and M.-M. Mac Low (2012). *ApJ* 756:62.

Mainardi, F. and G. Pagnini (2002). "Waves and Stability in Continuous Media".

Majda, A., J. McDonough, and S. Osher (1978). *Mathematics of Computation* 32:144:1041–1081.

Mamatsashvili, G.R. (2011). PhD thesis. University of Edinburgh.

Mann, R.K., S.M. Andrews, J.A. Eisner, J.P. Williams, M.R. Meyer, J. Di Francesco, J.M. Carpenter, and D. Johnstone (2015). *ApJ* 802:77.

Marconi, A. and L.K. Hunt (2003). *ApJ* 589:L21.

Marois, C., B. Macintosh, T. Barman, B. Zuckerman, I. Song, J. Patience, D. Lafreniére, and R. Doyon (2008). *Science* 322:1348.

Mayer, L., G. Lufkin, T. Quinn, and J. Wadsley (2007). *ApJ* 661:L77.

Meheut, H., F. Casse, P. Varniere, and M. Tagger (2010). *A&A* 516:A31.

Meheut, H., C. Yu, and D. Lai (2012b). *MNRAS* 422:2399–2406.

Meheut, H., R. Keppens, F. Casse, and W. Benz (2012a). *A&A* 542:A9.

Meheut, H., R.V.E. Lovelace, and D. Lai (2013). *MNRAS* 430:1988–1993.

Mejía, A.C., R.H. Durisen, M.K. Pickett, and K. Cai (2005). *ApJ* 619:1098.

Merloni, A. (2003). *MNRAS* 341:1051.

Meszaros, P., W. Nagel, and J. Ventura (1980). *ApJ* 238:1066–1080.

Michikoshi, S., A. Fujii, E. Kokubo, and H. Salo (2015). *ApJ* 812:151.

22nd Texas Symposium on Relativistic Astrophysics (2005).

Naoz, S., T. Fragos, A. Geller, A.P. Stephan, and F.A. Rasio (2016). *ApJL* 822:L24.

Narayan, R., T. Piran, and P. Kumar (2001). *ApJ* 557:949.

Nayakshin, S. (2010). *MNRAS* 408:L36.

Nelson, A.F., W. Benz, and T.V. Ruzmaikina (2000). *ApJ* 529:357.

Nelson, R.P., O. Gressel, and O.M. Umurhan (2013). *MNRAS* 435:2610–2632.

Osaki, Y. (1974). *PASJ* 26:429.

Ostriker, J.P. and K. Davidson (1973). *X- and Gamma-Ray Astronomy - Volume 55 of International Astronomical Union Symposia*. Ed. by H. Bradt and R. Giacconi. Springer Science & Business Media.

Owen, J.E., B. Ercolano, C.J. Clarke, and R.D. Alexander (2010). *MNRAS* 401:1415–28.

Paardekooper, S.-J. (2012). *MNRAS* 421:3286–3299.

Papaloizou, J.C.B. and D.N.C. Lin (1988). *ApJ* 331:838.

Papaloizou, J.C.B. and D.N.C. Lin (1989). *ApJ* 344:645.

Papaloizou, J.C.B. and R.P. Nelson (2003). *MNRAS* 339:983.

Papaloizou, J.C.B. and J.E. Pringle (1985). *MNRAS* 213:799.

Papaloizou, J.C.B. and J.E. Pringle (1987). *MNRAS* 225:267.

Papaloizou, J.C.B. and G.J. Savonije (1991). *MNRAS* 248:353.

Papaloizou, J.C.B. and G.Q.G. Stanley (1986). *MNRAS* 220:593.

Parikh, A., J. Jose, and G. Sala (2014). *AIP Advances* 4.

Patterson, J. (1984). *ApJS* 54.

Pérez, L.M., J.M. Carpenter, S.M. Andrews, L. Ricci, A. Isella, H. Linz, A.I. Sargent, D.J. Wilner, T. Henning, A.T. Deller, C.J. Chandler, C.P. Dullemond, J. Lazio, K.M. Menten, S.A. Corder, S. Storm, L. Testi, M. Tazzari, W. Kwon, N. Calvet, J.S. Greaves, R.J. Harris, and L.G. Mundy (2016). *Science* 353:1519–1521.

Petersen, M.R., K. Julien, and G.R. Stewart (2007a). *ApJ* 658:1236.

Petersen, M.R., G.R. Stewart, and K. Julien (2007b). *ApJ* 658:1252.

Pfalzner, S. (2008). *Astron. AStrophys.* 492:735.

Piran, T. (1978). *ApJ* 221:652.

Piran, T. (2004). *Reviews of Modern Physics* 76:1143–1210.

Prendergast, K.H. and G.R. Burbidge (1968). *ApJ* 151.

Prialnik, D. (1986). *ApJ* 310:222–237.

Pringle, J.E. (1976). *MNRAS* 177:65.

Pringle, J.E. (1981). *ARA&A* 19:137–162.

Pringle, J.E. and M.J. Rees (1972). *Astron. & Astrophys.* 21:1–9.

Pringle, J.E., M.J. Rees, and A.G. Pacholczyk (1973). *A&A* 29:179–184.

Rafikov, R.R. (2005). *ApJ* 621:L69.

Rafikov, R.R. (2007). *ApJ* 662:642.

Rayleigh, Lord (1880). *Proc. Lond. Math. Soc.* 14:170–177.

Reshotko, E. (2001). *Physics of Fluids* 13:1067.

Riaz, B., G. Lodato, D. Stamatellos, and J.E. Gizis (2012). *MNRAS* 422:L6.

Rice, W.K.M., P.J. Armitage, M.R. Bate, and I.A. Bonell (2003). *MNRAS* 339:1025.

Rice, W.K.M., G. Lodato, and P.J. Armitage (2005). *MNRAS* 364:L56–L60.

Rice, W.K.M., P.J. Armitage, G.R. Mamatsashvili, G. Lodato, and C.J. Clarke (2011). *MNRAS* 418:1356–1362.

Richard, D. (2001). PhD thesis. Univ. Paris VII.

Rincon, F., G.I. Ogilvie, and C. Cossu (2007). *A&A* 463:817.

Robinson, E.L. (1976). *Annu. Rev. Astron. Astrophys.*

Romanova, M.M, G.V. Ustyugova, A.V. Koldoba, and R.V.E. Lovelace (2004). *ApJ* 610:920–932.

Ruden, S.P. and D.N.C. Lin (1986). *ApJ* 308:883.

Ruden, S.P. and J.B. Pollack (1991). *ApJ* 375:740.

Ruden, S.P., J.C.B. Papaloizou, and D.N.C. Lin (1988). *ApJ* 329:739.

Rüdiger, G., R. Tschäpe, and L.L. Kitchatinov (2002). *MNRAS* 332:435.

Sakimoto, P.J. and F.V. Coroniti (1981). *ApJ* 247:19.

Salo, H. (1992). *Nature* 359:619–621.

Salpeter, E.E. (1964). *ApJ* 140:796–800.

Sano, T. and J.M. Stone (2002). *ApJ* 577:534.

Sari, R. and P. Goldreich (2004). *ApJ* 606:L77–L80.

Saslaw, W.C. (1968). *MNRAS* 138:337–358.

Schmid, P.J. (2000). *Physics of Plasmas* 7:1788.

Schmit, U. and W.M. Tscharnuter (1999). *Icarus* 138:173.

Sellwood, J.A. and R.G. Carlberg (1984). *ApJ* 282:61.

Shakura, N.I. and R.A. Sunyaev (1973). *A&A* 24:337–355.

Shakura, N.I. and R.A. Sunyaev (1976). *MNRAS* 175:613.

Shen, Y., J.M. Stone, and T.A. Gardiner (2006). *ApJ* 653:513–524.

Shi, J.-M., Z. Zhu, J.M. Stones, and E. Chiang (2016). *MNRAS* 459:982–998.

Shu, F.H., R.J. Allen, S. Lizano, and D. Galli (2007). *ApJ* 662:L75.

Simon, J.B., K. Beckwith, and P.J. Armitage (2012). *MNRAS* 422:2685–2700.

Sirko, E. and J. Goodman (2003). *MNRAS* 341:501.

Stamatellos, D. and A.P. Whitworth (2009). *MNRAS* 392:413–427.

Stamatellos, D., A.P. Whitworth, and D.A. Hubber (2012). *MNRAS* 427:1182–1193.

Stella, L. and R. Rosner (1984). *ApJ* 277:312.

Stoll, M.H.R. and W. Kley (2014). *A&A* 572:A77.

Stone, J.M., J.F. Hawley, C.F. Gammie, and S.A. Balbus (1996). *ApJ* 463:656–673.

Streett, C.L. (1987). *International Journal for Numerical Methods In Fluids* 7:1159–1189.

Sun, Z.-P., G. Schubert, and G.A. Glatzmaier (1993). *Science* 260:661.

Szilágyi, B., L. Lindblom, and M.A. Scheel (2009). *Phys. Rev.* D80:124010.

Tagger, M. (2001). *A&A* 380:750.

Taylor, G.I. (1936). *Royal Society of London Proceedings Series A* 157:546.

Taylor, M. (1994). *Journal of Computational Physics* 110:407–418.

Terebey, S., F.H. Shu, and P. Cassen (1984). *ApJ* 286:529–551.

Testi, L., T. Birnstiel, L. Ricci, S. Andrews, J. Blum, J. Carpenter, C. Dominik, A. Isella, A. Natta, J.P. Williams, and D.J. Wilner (2014). *Protostars and Planets VI*. Ed. by H. Beuther, R.S. Klessen, C.P. Dullemond, and T. Henning. The University of Arizona Press.

Thompson, T.A., E. Quataert, and N. Murray (2005). *ApJ* 630:167.

Tichy, W. (2006). *Phys. Rev.* D74:084005.

Tillmark, N. and P. Alfredsson (1996). *Advances in Turbulence VI*. Ed. by S. Gavrilakis, L. Machiels, and P. Monkewitz, p. 391.

Tobin, J.J., L. Hartmann, H.-F. Chiang, D.J. Wilner, L.W. Looney, L. Loinard, N. Calvet, and P. D'Alessio (2013). *ApJ* 771:48.

Tobin, J.J., L.W. Looney, D.J. Wilner, W. Kwon, C.J. Chandler, T.L. Bourke, L. Loinard, H.-F. Chiang, S. Schnee, and X. Chen (2015). *ApJ* 805:125.

Toomre, A. (1964). *ApJ* 139:1217–1238.

Toomre, A. (1981). "Structure and Evolution of Normal Galaxies". Ed. by S.M. Fall and D. Lynden-Bell, 111–136.

Townsley, D.M. and B.T. Gansicke (2008).

Trefethen, L.N., A.E. Trefethen, S.C. Reddy, and T.A. Driscoll (1993). *Science* 261:578.

Trumper, J.E. and G. Hasinger (2008). *The Universe in X-rays*. Springer-Verlag Berlin Heidelberg.

Umurhan, O.M. and O. Regev (2004). *A&A* 427:855.

Vanon, R. and G.I. Ogilvie (2016). *MNRAS* 463:3725–3736.

Vanon, R. and G.I. Ogilvie (2017). *MNRAS* 466:2590–2601.

Varnière, P. and M. Tagger (2006). *A&A* 446:L13–L16.

Velikhov, E.P. (1959). *J. Exp. Theor. Phys. (USSR)* 36:1398.

Von Weizsäcker, C.F. (1948). *Z. Naturforsch* 3a:524–539.
Vorobyov, E.I. and S. Basu (2005). *ApJ* 633:L137.
Vorobyov, E.I. and S. Basu (2006). *ApJ* 650:956.
Vorobyov, E.I. and S. Basu (2010). *ApJ* 719:1896.
Waleffe, F. (1997). *Physics of Fluids* 9:883.
Waltz, R.E., J.M. Candy, and M.N. Rosenbluth (2002). *Phys. Plasmas* 9:1938.
Wang, Y.-M. and G.L. Welter (1981). *Astron. Astrophys.* 102:97–102.
Warner, B. and W.L. Peters (1972). *Mon. Not. R. astr. Soc.* 160.
Webbink, R.F., M. Livio, J.W. Truran, and M. Orio (1987). *ApJ* 314:653–672.
Weidenschilling, S.J. (1977). *MNRAS* 180:157.
Wendt, G. (1933). *Ing. Arch.* 4:577–595.
Whipple, F.L. (1972). *Proc. 21st Nobel Symp., From Plasma to Planet*. Ed. by A. Elvius. New York: Wiley, p. 211.
Whitworth, A.P. and D. Stamatellos (2006). *A&A* 458:817–829.
Willerding, E. (1992). *EM&P* 56:173–192.
Williams, G.P. (1985). *Adv. Geophys.* 28A:381–349.
Williams, J.P. and L.A. Cieza (2011). *Annu. Rev. Astro. Astrophys.* 49:67–117.
Yecko, P.A. (2004). *A&A* 425:385–393.
Yungelson, L., M. Livio, and J.W. Truran (1996). *ApJ* 466.
Zeldovich, Ya.B. (1964). *Soviet Physics Doklady* 9:195.
Zinn, P.-C., E. Middelberg, R.P. Norris, and R.-J. Dettmar (2013). *ApJ* 774.